

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:31:37 ; Search time 59 Seconds
(without alignments)
1805.431 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEL.....PGPMDCHRALEPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		8						
Result		Query						
No.	Score	Match	Length	DB	ID		Description	
1	1979	100.0	377	5	AAE28936	Aae28936	Human sod	
2	1970	99.5	377	6	'AAE29906	Aae29906	Human tra	
3	1180	59.6	325	5	ABG76899	Abg76899	Human ile	
4	884	44.7	348	2	AAR77224	Aar77224	Hamster i	
5	860.5	43.5	348	2	AAR77225	Aar77225	Human ile	
6	860.5	43.5	348	6	AAO19649	Aao19649	Human ile	
7	559.5	28.3	362	7	ADD48705	Add48705	Rat Prote	
8	553	27.9	349	6	AAE37351	Aae37351	Human sod	
9	553	27.9	349	7	ADD48707	Add48707	Human Pro	

10	393.5	19.9	270	6	ABP75825	Abp75825	Human	sec
11	386.5	19.5	491	5	AAE13283	Aae13283	Human	tra
12	381	19.3	491	5	ABP43962	Abp43962	clone	IMA
13	380.5	19.2	490	6	ABU69595	Abu69595	Human	NF-
14	380.5	19.2	490	6	ABU69621	Abu69621	Human	NF-
15	363.5	18.4	225	5	AAE21252	Aae21252	Human	gen
16	354.5	17.9	454	4	ABG00575	Abg00575	Novel	hum
17	322.5	16.3	367	5	ABP70059	Abp70059	Human	NOV
18	322.5	16.3	367	5	ABP70061	Abp70061	Human	NOV
19	322.5	16.3	438	5	AAE28937	Aae28937	Human	sod
20	322.5	16.3	438	5	ABP70060	Abp70060	Human	NOV
21	322.5	16.3	438	6	AAE29908	Aae29908	Human	tra
22	321.5	16.2	202	6	ADA56762	Ada56762	Human	sec
23	321.5	16.2	202	6	ADA40611	Ada40611	Human	sec
24	321.5	16.2	202	7	ADC74028	Adc74028	Human	sec
25	318.5	16.1	207	5	AAE21253	Aae21253	Human	gen
26	298.5	15.1	319	6	ABM72920	Abm72920	Staphyloc	
27	290	14.7	345	7	ABM73656	Abm73656	DNA	clone
28	284	14.4	338	5	ABB91897	Abb91897	Herbicida	
29	280.5	14.2	325	6	ADA33813	Ada33813	Acinetoba	
30	280	14.1	356	3	AAG22453	Aag22453	Arabidops	
31	273	13.8	423	3	AAG48158	Aag48158	Arabidops	
32	268.5	13.6	335	4	AAG91138	Aag91138	C glutami	
33	266	13.4	455	4	ABB70896	Abb70896	Drosophil	
34	265.5	13.4	196	5	AAE21198	Aae21198	Human	gen
35	265.5	13.4	196	5	ABG64873	Abg64873	Human	alb
36	265.5	13.4	196	6	ADA57328	Ada57328	Human	sec
37	265.5	13.4	196	6	ADA41200	Ada41200	Human	sec
38	265.5	13.4	196	7	ADC74403	Adc74403	Human	sec
39	264.5	13.4	324	4	AAG92300	Aag92300	C glutami	
40	264	13.3	271	3	AAG22454	Aag22454	Arabidops	
41	257	13.0	285	3	AAG48159	Aag48159	Arabidops	
42	255	12.9	268	3	AAG22455	Aag22455	Arabidops	
43	250	12.6	607	4	ABG00574	Abg00574	Novel	hum
44	248	12.5	282	3	AAG48160	Aag48160	Arabidops	
45	248	12.5	353	5	ABP65901	Abp65901	Bifidobac	

ALIGNMENTS

RESULT 1

AAE28936

ID AAE28936 standard; protein; 377 AA.

XX

AC AAE28936;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human sodium/bile-like transporter protein #1.

XX

KW Human; sodium/bile-like transporter; novel human protein; drug screening;

KW NHP; cancer; cosmetic; nutraceutical; gene therapy; cytostatic;

KW chromosome 4.

XX

OS Homo sapiens.

XX

PN WO200272774-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-MAR-2002; 2002WO-US007438.
 XX
 PR 12-MAR-2001; 2001US-0275009P.
 PR 17-APR-2001; 2001US-0284152P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Wilganowski NL, Nepomnichy B, Burnett MB, Hu Y;
 XX
 DR WPI; 2002-723334/78.
 DR N-PSDB; AAD46333.
 XX
 PT New protein and nucleic acid molecule, useful for diagnosing or treating
 PT diseases, e.g. cancer, for drug screening, clinical trial monitoring,
 PT pharmacogenomics, and for cosmetic or nutraceutical applications.
 XX
 PS Claim 4; Page 37-38; 41pp; English.
 XX
 CC The invention relates to novel human proteins (NHP), sodium/bile-like
 CC transporter and their nucleic acids. The invention is useful for
 CC identifying the protein which may be used for diagnosis, clinical trial
 CC monitoring, drug screening, pharmacogenomics, treatment of diseases such
 CC as cancer, and for cosmetic or nutraceutical applications. The nucleic
 CC acid molecule may also be used as hybridisation probes for screening
 CC libraries, assessing gene expression patterns, and in amplification
 CC assays. It is also used in gene therapy. The present sequence is human
 CC sodium/bile-like transporter protein. The gene encoding this protein is
 CC located at chromosome 4
 XX
 SQ Sequence 377 AA;

Query Match 100.0%; Score 1979; DB 5; Length 377;
 Best Local Similarity 100.0%; Pred. No. 3e-205;
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWS	60
Db	1	MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWS	60
Qy	61	HIRRPWGIAGVLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD	120
Db	61	HIRRPWGIAGVLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD	120
Qy	121	GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGV	180
Db	121	GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGV	180
Qy	181	YVNYRWPQKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT	240
Db	181	YVNYRWPQKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT	240
Qy	241	GFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSF TAEHLVQMLS FPLAYGLFQLID	300

```

Db      241 GFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLID 300
Qy      301 GFLIVAAYQTYKRRLKNKHGKKNKGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GFLIVAAYQTYKRRLKNKHGKKNKGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
Qy      361 MDCHRALEPVGHITSCE 377
        ||||||||||||||||
Db      361 MDCHRALEPVGHITSCE 377

```

RESULT 2

AAE29906

ID AAE29906 standard; protein; 377 AA.

XX

AC AAE29906;

XX

DT 24-FEB-2003 (first entry)

XX

DE Human transporter and ion channel (TRICH) protein #6.

XX

KW Human; transporter and ion channel; TRICH; neurodegenerative disorder;
 KW Parkinson's disease; Alzheimer's disease; muscular disorder; transgenic;
 KW myotonic dystrophy; catatonia; endocrine disorder; diabetes; cytostatic;
 KW Grave's disease; cancer; leukaemia; cervical; immunological; scleroderma;
 KW systemic lupus erythematosus; allergy; gastrointestinal; Crohn's disease;
 KW Goodpasture's syndrome; infection; cardiovascular; fungicide; nootropic;
 KW hepatic disease; cirrhosis; gene therapy; uropathic; anti-HIV; virucide;
 KW atherosclerosis; antiparasitic; protozoacide; antibacterial.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	Domain	28. .56
FT		/note= "Transmembrane domain"
FT	Domain	39. .220
FT		/note= "Sodium bile acid symporter domain"
FT	Peptide	41. .97
FT		/label= Signal_peptide
FT	Domain	69. .89
FT		/note= "Transmembrane domain"
FT	Domain	95. .115
FT		/note= "Transmembrane domain"
FT	Protein	98. .377
FT		/note= "Human mature TRICH protein"
FT	Domain	131. .153
FT		/note= "Transmembrane domain"
FT	Domain	159. .182
FT		/note= "Transmembrane domain"
FT	Domain	191. .218
FT		/note= "Transmembrane domain"
FT	Domain	220. .248
FT		/note= "Transmembrane domain"

XX

PN WO200277237-A2.

XX

PD 03-OCT-2002.

```

XX      08-FEB-2002; 2002WO-US003657.
XX
PR      09-FEB-2001; 2001US-0267892P.
PR      23-FEB-2001; 2001US-0271168P.
PR      02-MAR-2001; 2001US-0272890P.
PR      16-MAR-2001; 2001US-0276860P.
PR      23-MAR-2001; 2001US-0278255P.
PR      30-MAR-2001; 2001US-0280538P.
PR      25-JAN-2002; 2002US-0351359P.
XX
PA      (INCY-) INCYTE GENOMICS INC.
XX
PI      Lee EA, Ding L, Baughn MR, Tribouley CM, Bruns CM, Elliott VS;
PI      Walia NK, Forsythe IJ, Raumann BE, Burford N, Lal PG, Thornton M;
PI      Gandhi AR, Arvizu C, Yao MG, Yue H, Xu Y, Hafalia AJA, Ison CH;
PI      Chen H;
XX
DR      WPI; 2003-018931/01.
DR      N-PSDB; AAD47353.
XX
PT      New TRICH polypeptides, useful for diagnosing, preventing, and treating
PT      disorders associated with an abnormal expression or activity of TRICH,
PT      e.g. neuromuscular, immunological, cardiovascular disorders, cancer and
PT      infection.
XX
PS      Claim 1; Page 158-159; 214pp; English.
XX
CC      The invention relates to human transporters and ion channels (TRICH) and
CC      their nucleic acids. The sequences of the invention are useful in
CC      diagnosing, preventing, and treating disorders associated with an
CC      abnormal expression or activity of TRICH, such as neurodegenerative
CC      disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular
CC      disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g.
CC      diabetes, Grave's disease), cancers (e.g. leukaemia, cervical or breast
CC      cancers), immunological disorders (e.g. scleroderma, systemic lupus
CC      erythematosus, allergies), gastrointestinal disorders (e.g. Crohn's
CC      disease), renal disorders (e.g. Goodpasture's syndrome), infections (e.g.
CC      viral, bacterial, fungal, parasitic, protozoal, helminthic),
CC      cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases
CC      (e.g. cirrhosis). TRICH or its fragments may also be used in screening
CC      for compounds that specifically bind to and modulate its activity. TRICH
CC      DNA can be used to create humanised animals or transgenic animals to
CC      model human disease. It is also used in gene therapy. The present
CC      sequence is human TRICH protein
XX
SQ      Sequence 377 AA;

Query Match          99.5%;  Score 1970;  DB 6;  Length 377;
Best Local Similarity 99.5%;  Pred. No. 2.8e-204;
Matches 375;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      1 MRANCSSSSACPANSSEEELPVGLEVHGNLELVFTVVTVMGLLMFSLGCSVEIRKLWS 60
        ||||||||||||||||||||| ||||||||||| ||||||||||| |||||||||||
Db      1 MRANCSSSSACPANSSEEELPVGLEAHGNLELVFTVVPVTVMGLLMFSLGCSVEIRKLWS 60

Qy      61 HIRRPWGIAVGLLCCQFGLMPFTAYLLAISFSLKPQVQAI AVLIMGCCPGGTISNIFTFWVD 120

```

```

      |||
Db      61 HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD 120
      |||
Qy      121 GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGV 180
      |||
Db      121 GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGV 180
      |||
Qy      181 YVNYRWPKQSKIILKIGAVVGGVLLL VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT 240
      |||
Db      181 YVNYRWPKQSKIILKIGAVVGGVLLL VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT 240
      |||
Qy      241 GFL LALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLID 300
      |||
Db      241 GFL LALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLID 300
      |||
Qy      301 GFLIVAAYQTYKRRLKNKHGKKN SGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
      |||
Db      301 GFLIVAAYQTYKRRLKNKHGKKN SGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
      |||
Qy      361 MDCHRALEPVGHITSCE 377
      |||
Db      361 MDCHRALEPVGHITSCE 377

```

RESULT 3

ABG76899

ID ABG76899 standard; protein; 325 AA.

XX

AC ABG76899;

XX

DT 05-NOV-2002 (first entry)

XX

DE Human ileal sodium/bile acid cotransporter-like protein.

XX

KW Human; NOVX; cardiomyopathy; atherosclerosis; cell signal processing;
KW breast cancer; Alzheimer's disease; epilepsy; Huntington's disease;
KW anxiety; behavioural disorder; multiple sclerosis; myasthenia gravis;
KW neurodegeneration; Parkinson's disease; pain; stroke; endometriosis;
KW autoimmune disease; allergy; addiction; asthma; transplantation;
KW graft versus host disease; systemic lupus erythematosus; scleroderma;
KW psoriasis; Crohn's disease; HIV infection; human immunodeficiency virus;
KW atherosclerosis; cirrhosis; rheumatoid arthritis; diabetes; pancreatitis;
KW thrombocytopenia; bleeding disorder; metabolic disorder; obesity;
KW glucose transport defect; glomerulonephritis; hypercalcaemia;
KW polycystic kidney disease; renal tubular acidosis; skin disorder;
KW congenital diarrhoea; respiratory disease; gastro-intestinal disease;
KW muscle disorder; bone disorder; joint disorder; skeletal disorder;
KW haematopoietic disorder; urinary system disorder; osteoporosis;
KW dental disease; dental infection; growth disorder; reproductive disorder;
KW hypogonadism; fertility disorder; viral infection; bacterial infection;
KW parasitic infection; metabolic pathway modulation; gene therapy;
KW zinc metalloprotease; ADAM-TS 7; alpha-2-macroglobulin precursor;
KW ileal sodium/bile acid cotransporter; prohibitin; MT; CIP4; spinesin;
KW macrophage stimulating protein precursor; fatty acid-binding protein;
KW gap junction beta-5 protein; hepsin/plasma transmembrane serine protease.

XX

OS Homo sapiens.

XX
PN WO200233087-A2.
XX
PD 25-APR-2002.
XX
PF 17-OCT-2001; 2001WO-US032496.
XX
PR 17-OCT-2000; 2000US-0241040P.
PR 17-OCT-2000; 2000US-0241058P.
PR 17-OCT-2000; 2000US-0241063P.
PR 17-OCT-2000; 2000US-0241243P.
PR 20-OCT-2000; 2000US-0242152P.
PR 23-OCT-2000; 2000US-0242482P.
PR 23-OCT-2000; 2000US-0242611P.
PR 23-OCT-2000; 2000US-0242612P.
PR 24-OCT-2000; 2000US-0242880P.
PR 24-OCT-2000; 2000US-0242881P.
PR 29-DEC-2000; 2000US-0259028P.
PR 20-FEB-2001; 2001US-0269813P.
PR 25-APR-2001; 2001US-0286324P.
PR 29-MAY-2001; 2001US-0294108P.
PR 09-JUL-2001; 2001US-0303698P.
PR 16-OCT-2001; 2001US-00981151.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Edinger S, Gerlach V, Macdougall JR, Malyankar UM, Smithson G;
PI Millet I, Peyman JA, Stone DJ, Gunther E, Ellerman K, Shimkets RA;
PI Padigar M, Guo X, Patturajan M, Taupier RJ, Burgess CE;
PI Zerhusen BD, Kekuda R, Spytek KA, Gangolli EA, Fernandes ER;
PI Gorman L;
XX
DR WPI; 2002-590434/63.
DR N-PSDB; ABS59328.
XX
PT Cytoplasmic, nuclear, membrane bound and secreted polypeptides and
PT nucleic acids encoding the polypeptides for diagnosing and treating e.g.
PT cancer, Alzheimer's disease, cardiomyopathy, metabolic disease and
PT diabetes.
XX
PS Claim 1; Page 50; 305pp; English.
XX
CC The present invention relates to new NOVX (NOV1-10) polypeptides. The
CC molecules of the invention are useful for treating or preventing a NOVX-
CC associated disorder, such as cardiomyopathy, atherosclerosis, or a
CC disorder related to cell signal processing and metabolic pathway
CC modulation in humans. NOVX polypeptides, nucleic acids and antibodies are
CC useful for treating or preventing disorders or syndromes including breast
CC cancer, Alzheimer's disease, epilepsy, Huntington's disease, anxiety,
CC behavioural disorders, multiple sclerosis, myasthenia gravis,
CC neurodegeneration, Parkinson's disease, pain, stroke, autoimmune disease,
CC allergies, addiction, asthma, endometriosis, graft versus host disease,
CC systemic lupus erythematosus, scleroderma, transplantation, psoriasis,
CC Crohn's disease, HIV (human immunodeficiency virus) infection,
CC atherosclerosis, cirrhosis, rheumatoid arthritis, diabetes,
CC thrombocytopenia, bleeding disorders, metabolic disorders, obesity,
CC glucose transport defect, glomerulonephritis, hypercalcaemia, polycystic


```

XX
PF 29-DEC-1994; 94WO-US014431.
XX
PR 29-DEC-1993; 93US-00176126.
XX
PA (UYWA-) UNIV WAKE FOREST.
XX
PI Dawson PA;
XX
DR WPI; 1995-246189/32.
DR N-PSDB; AAQ91108.
XX
PT Hamster and human ileal and bile acid transport DNA and protein - useful
PT in treatment of e.g. hypercholesterolaemia, diabetes and various
PT digestive diseases, and in gene therapy to restore bile acid uptake
PT activity.
XX
PS Claim 34; Page 104-106; 148pp; English.
XX
CC The ileal/renal bile acid cotransporter protein is useful in the
CC treatment of hypercholesterolaemia, diabetes, heart disease, liver
CC disease and various digestive disorders. The cDNA may be used in gene
CC therapy to restore bile acid uptake activity to patients whose ileum has
CC been surgically resected for diseases such as Crohn disease, patients
CC born with congenital defects in the bile transporter, and patients
CC suffering from adult-onset chronic idiopathic bile acid diarrhoea. The
CC DNA and protein may be used in screening methods as modulators of
CC ileal/renal bile acid cotransport activity
XX
SQ Sequence 348 AA;

Query Match          44.7%; Score 884; DB 2; Length 348;
Best Local Similarity 46.9%; Pred. No. 1.le-86;
Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | : : | : | | : | |::: |:|::|::|:: | | :
Db      3 NSSICNPNATICEGDSCIAPESNFNAILSVVMSTVLTLILLALVMFSGMCNVELHKFLGHL 62

Qy      63 RRPWGIAGVLLCQFGLMPFTAYLLAISFLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| | | ||||:| | ::::| : |||: | | ||||| | | :|||
Db      63 RRPWGI VVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCCPGGTASNILAYWVDGD 122

Qy      123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
      |||:|||||: | |||||:| | : | | :| | :| | |||: |::
Db      123 MDLSVSM T TCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPV SIGMYV 182

Qy      183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:|:::| |||||:: | :|:::| | | : : | : | | |:| : | |
Db      183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242

Qy      243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGF 302
      || | | |||:| | | | :| | :| |||: | | : :| | | :| :
Db      243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302

Qy      303 LIVAAYQTYKRRLLKNKHGKKN SGCTEVCHTRKS--TSSRETN AFLEVNEE 350
      :: | | | : | | | : | : | | : : | :

```

Db 303 ILLGAYVAYKK----CHGKNNTLQEKTDNEMEPRSSFQETNKGFPDEK 348

RESULT 5

AAR77225

ID AAR77225 standard; protein; 348 AA.

XX

AC AAR77225;

XX

DT 17-DEC-1995 (first entry)

XX

DE Human ileal/renal bile acid cotransporter.

XX

KW Ileal/renal bile acid cotransporter; therapeutic; gene therapy;

KW diagnostic.

XX

OS Homo sapiens.

XX

PN WO9517905-A1.

XX

PD 06-JUL-1995.

XX

PF 29-DEC-1994; 94WO-US014431.

XX

PR 29-DEC-1993; 93US-00176126.

XX

PA (UYWA-) UNIV WAKE FOREST.

XX

PI Dawson PA;

XX

DR WPI; 1995-246189/32.

DR N-PSDB; AAQ91109.

XX

PT Hamster and human ileal and bile acid transport DNA and protein - useful

PT in treatment of e.g. hypercholesterolaemia, diabetes and various

PT digestive diseases, and in gene therapy to restore bile acid uptake

PT activity.

XX

PS Claim 34; Page 111-114; 148pp; English.

XX

CC The ileal/renal bile acid cotransporter protein is useful in the
CC treatment of hypercholesterolaemia, diabetes, heart disease, liver
CC disease and various digestive disorders. The cDNA may be used in gene
CC therapy to restore bile acid uptake activity to patients whose ileum has
CC been surgically resected for diseases such as Crohn disease, patients
CC born with congenital defects in the bile transporter, and patients
CC suffering from adult-onset chronic idiopathic bile acid diarrhoea. The
CC DNA and protein may be used in screening methods as modulators of
CC ileal/renal bile acid cotransport activity

XX

SQ Sequence 348 AA;

Query Match 43.5%; Score 860.5; DB 2; Length 348;

Best Local Similarity 45.6%; Pred. No. 3.7e-84;

Matches 160; Conservative 68; Mismatches 104; Indels 19; Gaps 4;

Qy 5 CSSSSACPANSSEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR 64

Db	14	CSGASCVVPESNFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR	64
Qy	65	PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMD	124
Db	65	PWGICVGFLLCQFGIMPLTGFILSVAFDILPLQAVVLIIGCCPGGTASNILAYWVDGDMD	124
Qy	125	LSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNGITLVCLTIPVAFGVVNY	184
Db	125	LSVSMTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPVSGMFVNH	184
Qy	185	RWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL	244
Db	185	KWPQKAKIILKIGSIAGAILIVLIAVVGILYQSAWIIAPKLWIIGTIFPVAGYSLGFL	244
Qy	245	ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFL	304
Db	245	ARIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPEELNVVFTFPLIYSIFQLAFAAIF	304
Qy	305	VAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITP	355
Db	305	LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFQP	345

RESULT 6

AAO19649

ID AAO19649 standard; protein; 348 AA.

XX

AC AAO19649;

XX

DT 28-MAR-2003 (first entry)

XX

DE Human ileal sodium-dependent bile acid transporter protein.

XX

KW Human; ileal sodium-dependent bile acid transporter gene; SLC10A2; SNP;

KW single nucleotide polymorphism; chromosome 13q33; cardiant;

KW antiarteriosclerotic; antilipemic.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Misc-difference	65
----	-----------------	----

FT		/note= "optionally Leu depending on SNP present in gene"
----	--	--

FT	Misc-difference	98
----	-----------------	----

FT		/note= "optionally Ile depending on SNP present in gene"
----	--	--

FT	Misc-difference	159
----	-----------------	-----

FT		/note= "optionally Ile depending on SNP present in gene"
----	--	--

FT	Misc-difference	290
----	-----------------	-----

FT		/note= "optionally Ser depending on SNP present in gene"
----	--	--

FT	Misc-difference	296
----	-----------------	-----

FT		/note= "optionally Leu depending on SNP present in gene"
----	--	--

FT	Misc-difference	316
----	-----------------	-----

FT		/note= "optionally Glu depending on SNP present in gene"
----	--	--

XX

PN WO200283944-A2.

XX

PD 24-OCT-2002.

XX
PF 11-APR-2002; 2002WO-GB001681.
XX
PR 17-APR-2001; 2001GB-00009296.
PR 19-APR-2001; 2001US-0284530P.
XX
PA (ASTR) ASTRAZENECA AB.
PA (ASTR) ASTRAZENECA UK LTD.
XX
PI Morten JEN;
XX
DR WPI; 2003-046927/04.
XX
PT Diagnosing polymorphism in SLC10A2 in a human for assessing the
PT pharmacogenetics of a drug for treating cardiovascular and hyperlipidemic
PT conditions, by determining the status of the human by reference to
PT polymorphism in SLC10A2.
XX
PS Claim 10; Page 21; 21pp; English.
XX
CC The present invention relates to a method of diagnosing polymorphisms in
CC SLC10A2 (human ileal sodium-dependent bile acid transporter gene) in a
CC human, which involves determining the status of the human by reference to
CC polymorphisms in SLC10A2. The method is useful for assessing the
CC pharmacogenetics of a drug acting at SLC10A2. The SLC10A2 gene
CC polymorphism is useful as a genetic marker in a linkage study. SLC10A2
CC drugs are also useful for treating cardiovascular (e.g. atherosclerosis)
CC and hyperlipidemic conditions. The SLC10A2 gene is found at chromosome
CC 13q33. The present sequence is the protein of the invention with
CC associated alternative amino acids
XX
SQ Sequence 348 AA;

Qy	5	CSSSSACPANSSSEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR	64
Db	14	CSGASCVPVPESNFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR	64
Qy	65	PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMD	124
Db	65	PWGICVGFCLCQFGIMPLTGFI LSVAFDILPLQAVVVLIIGCCPGGTASNILAYWVDGDMD	124
Qy	125	LSISMTTCSTVAALGMMPLCIYLTWSWSLQQNLTIPYQNIGITLVCLTIPVAFGVVNY	184
Db	125	LSVSMTTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPV SIGMFVNH	184
Qy	185	RWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL	244
Db	185	KWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWIIAPKLWII GTIFPVAGYSLGFL	244
Qy	245	ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLS FPLAYGLFQLIDGFLI	304
Db	245	ARIAGLPWYRCRTVAFETGMONTOLCSTIVQLSFTPEELNVVETFPLIYSIFOLAFAAIF	304

Qy 305 VAA YQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITP 355
 : | ||: ||| : | : :| : | |
 Db 305 LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFQP 345

RESULT 7

ADD48705

ID ADD48705 standard; protein; 362 AA.

XX

AC ADD48705;

XX

DT 29-JAN-2004 (first entry)

XX

DE Rat Protein P26435, SEQ ID NO 14414.

XX

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX

OS Rattus norvegicus.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; P26435.

XX

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a

KW arteriosclerosis; coronary artery ligation; rheumatic heart disease;
 KW heart failure; hypertension; cardiomyopathy; myocardial infarction;
 KW arterial inflammation; microembolism; atherosclerosis; endocarditis;
 KW vascular heart disease; valvular disease; arrhythmia; gene therapy;
 KW sinus node dysfunction; sodium-bile acid cotransporter.
 XX
 OS Homo sapiens.
 XX
 PN WO2003039341-A2.
 XX
 PD 15-MAY-2003.
 XX
 PF 05-NOV-2002; 2002WO-US035538.
 XX
 PR 05-NOV-2001; 2001US-0339582P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Logan TJ, Chun M, Galvin KM;
 XX
 DR WPI; 2003-441437/41.
 DR N-PSDB; AAD56518.
 XX
 PT Treating a subject having a cardiovascular disorder, e.g. angina,
 PT arrhythmia, or restenosis, comprises administering a 139, 258, 1261,
 PT 1486, 2398, 2414, 7660, 8587, 10183, 10550, 12680, 17921, 32248, 60489 or
 PT 93804 modulator.
 XX
 PS Disclosure; Page 109-110; 124pp; English.
 XX
 CC The invention relates to methods and compositions for treating a subject
 CC having a cardiovascular disorder using 139, 258, 1261, 1486, 2398, 2414,
 CC 7660, 8587, 10183, 10550, 12680, 17921, 32248, 60489 or 93804 modulator.
 CC The invention is useful for treating a cardiovascular disorder, including
 CC arteriosclerosis, atherosclerosis, vascular wall remodeling, restenosis,
 CC cardiac hypertrophy, ischaemia reperfusion injury, arterial inflammation,
 CC ventricular remodelling, rapid ventricular pacing, tachycardia, coronary
 CC microembolism, bradycardia, pressure overload, aortic bending, coronary
 CC artery ligation, vascular heart disease, valvular disease, including but
 CC not limited to, valvular degeneration caused by calcification, rheumatic
 CC heart disease, endocarditis, or complications of artificial valves;
 CC atrial fibrillation, long-QT syndrome, congestive heart failure, sinus
 CC node dysfunction, angina, heart failure, hypertension, atrial flutter,
 CC atrial fibrillation, pericardial disease, including but not limited to
 CC pericardial effusion and pericarditis, cardiomyopathies (e.g. dilated
 CC cardiomyopathy or idiopathic cardiomyopathy), myocardial infarction,
 CC coronary artery disease, coronary artery spasm, ischaemic disease,
 CC arrhythmia, sudden cardiac death, and cardiovascular developmental
 CC disorders. The invention is also useful in gene therapy. The present
 CC sequence is human sodium/bile acid cotransporter protein. This sequence
 CC is used to illustrate the method of the invention
 XX
 SQ Sequence 349 AA;

Query Match 27.9%; Score 553; DB 6; Length 349;
 Best Local Similarity 36.0%; Pred. No. 7.6e-51;
 Matches 124; Conservative 77; Mismatches 109; Indels 34; Gaps 10;

Qy 31 ELVFTTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISEF 90
 :| :|: |: :| |||:| |: :|: |: |:|:| |:|:| |
 Db 24 DLALSVILVFMLFFIMLSLGCTMEFSKIKAHLWKPKGLAIALVAQYGIMPLTAFVLGKVF 83

Qy 91 SLKPVQAIAVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYT- 149
 || :|:|:|:| || ||| :|:|:| : |||:| ||||| ||||| :|:|:
 Db 84 RLKNIEALAILVCGCSPGGNLSNVFSLAMKGDMLNSIVMTTCSTFCALGMMPLLLYYSR 143

Qy 150 --WSWSLQQNLTIPIYQNIIGITLVCLTIPVAFGVVYNYRWPKQSKIILKIGAVVGGVLLLV 207
 : |: :|:| |:| |:| |:| :|:| :|:| :|:| :|:| :|:
 Db 144 GIYDGDLDKDK--VPYKGIVISLVLVLIPICTIGIVLKSRRPQYMYVIKGGMII----ILL 197

Qy 208 VAVAGVVLAKGSWNSDI-----TLTISFIFPLIGHVTGFL-ALFTHQSWQRC-RTIS 259
 :|| ||: : | |:| |:| |:| :|:| ||| ||| :|:|
 Db 198 CSVAVTVLSAINVGKSIMFAMTPLLITSSLMPFIFGLLGYVLSALFCLNG--RCRRTVS 255

Qy 260 LETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGFLIVAAYQTYKRRLKNKH 319
 :||| ||:|:| |:| :|:| | : : ||| |:| |:| :| |:| :| :|:| :| :|:
 Db 256 METGCQNVQLCSTILNVAFPPEVIGLPLFFFPLLYMIFQLGEGLLLIAIFWCYE-KFKTPK 314

Qy 320 GKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGPMDC 363
 | |:| :| :| | ||: | ||
 Db 315 DK-----TKMIYTAATT-----EETIPGALNGTYKGEDC 344

RESULT 9

ADD48707

ID ADD48707 standard; protein; 349 AA.

XX

AC ADD48707;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human Protein Q14973, SEQ ID NO 14416.

XX

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX

OS Homo sapiens.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.
 DR GENBANK; Q14973.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 349 AA;

Query Match 27.9%; Score 553; DB 7; Length 349;
 Best Local Similarity 36.0%; Pred. No. 7.6e-51;
 Matches 124; Conservative 77; Mismatches 109; Indels 34; Gaps 10;

Qy 31 ELVFTTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISF 90
 :| :|: |: :| |||:| |: |:| :| |:| :| |:|:| |:|:| |
 Db 24 DLALSUILVFMLEFFIMLSLGCTMEFSKIKAHLLWKPKGLAIALVAQYGYIMPLTAEVLGKVF 83
 Qy 91 SLKPVQAI AVLIMGCCPGGTISNIFTFVWDGDMDL SISMTTCSTVAALGMMPLCIYLYT- 149
 || ::|:|:|: || ||| :|:|: : |||:| ||||| ||||| :|:|:
 Db 84 RLKNIEALAILVCGCSPGGNLSNVFSLAMKGMNLSIVMTTCSTFCALGMMPLLLYIYSR 143
 Qy 150 --WSWSLQQNLTIPIYQNI GITLVCLTIPVAFGVYVNYRWPKQSKIILKIGAVVGGVLLLV 207
 : |: :|:| | |:| : || |: : : |: : :| | : :|:
 Db 144 GIYDGLDKD--VPYKGIVISLVVLIPCTIGIVLKS KRPQYMRVVIKGGMII----ILL 197
 Qy 208 VAVAGVVLAKGSWN SDI-----TLLTISFIFPLIGHVTGFLL-ALFTHQSWQRC-RTIS 259
 :|| ||: : | |: | : | || : |:| ||| || ||:|
 Db 198 CSVAVTVLSAINVGK SIMFAMTPLL IATSSLM PFI GFL LGYVLSALFCLNG--RCRRTVS 255

QY 260 LETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKRRLKNKH 319
 :||| ||:|:| |:| ::| | : : ||| | :||| :| |::| : |: : |
 Db 256 METGCQNVQLCSTILNVAFFPEVIGPLFFFPLLYMIFQLGEGLLLIAIFWCYE-KFKTPK 314

QY 320 GKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGPMDC 363
 | |:: :| :| | ||: | ||
 Db 315 DK-----TKMIYTAATT-----EETIPGALGNGTYKGEDC 344

RESULT 10

ABP75825

ID ABP75825 standard; protein; 270 AA.

XX

AC ABP75825;

XX

DT 10-FEB-2003 (first entry)

XX

DE Human secretory polypeptide SPTM SEQ ID NO 1009.

XX

KW Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;

KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;

KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;

KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;

KW anti-inflammatory; immunosuppressive; neuroprotective; nootropic;

KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;

KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;

KW secretory polynucleotide; secretory protein.

XX

OS Homo sapiens.

XX

PN WO200283876-A2.

XX

PD 24-OCT-2002.

XX

PF 27-MAR-2002; 2002WO-US009921.

XX

PR 29-MAR-2001; 2001US-0280067P.

PR 29-MAR-2001; 2001US-0280068P.

PR 16-MAY-2001; 2001US-0291280P.

PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.

PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.

PR 20-JUN-2001; 2001US-0300001P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;

PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX

DR WPI; 2003-075543/07.

DR N-PSDB; ABZ36267.

XX

PT New human secretory proteins and polynucleotides, useful for diagnosing,
PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
PT neurological disorders (e.g. Alzheimer's), or cell proliferations or
PT cancers.

XX

PS Claim 27; SEQ ID NO 1009; 458pp + Sequence Listing; English.

XX

CC The invention relates to a secretory polynucleotide (designated sptm)
CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
CC naturally occurring polynucleotide sequence at least 90 % identical to
CC the polynucleotide sequence, a polynucleotide complementary to them or an
CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
CC treating, preventing or diagnosing a disease or condition associated with
CC the expression of functional SPTM. These are particularly useful for
CC diagnosing, treating or preventing autoimmune/inflammatory disorders
CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
CC breast, cervix or prostate). The present sequence is one of the SPTM
CC proteins of the invention (ABP75384-ABP75962). Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 270 AA;

Query Match 19.9%; Score 393.5; DB 6; Length 270;
Best Local Similarity 34.2%; Pred. No. 1e-33;
Matches 82; Conservative 47; Mismatches 76; Indels 35; Gaps 4;

Qy 98 IAVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSW---SL 154
:| | | : | | | | | : | | : : | | | | : | | | | | : | | : | | | : | | :
Db 2 VAVLLCGCCPGGNLSNLSLLVDGMDNLSIIMTISSTLLALVLMPLCLWIYSWAWINTPI 61

Qy 155 QQNLTIPIYQNIIGITLVCLTIPVAFGVYVNYRWPQSKIILKI-----GA 198
| : | : : | | : | | : | : : : : | : | : |
Db 62 VQ--LLPLGTVTLTLCSTLIPIGLGVFIYKYSRVADYIVKVSLSLLVTLVVLFIIMTGT 119

Qy 199 VVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTI 258
: : | | : : | : | : | : : : | : : : : | : : : : | : : : : | : : : :
Db 120 MLGPELLASIPAAVYVIA-----IFMPLAGYASGYGLATLFHLPNCKRTV 165

Qy 259 SLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGFLIVAAYQTYKRRLKNK 318
| | | : | : | : : | : : : : | : | : | : : | : : | : : | : : | : : |
Db 166 CLETGSQNVQLCTAILKLAFFPPQFIGSMYMFLLYALFQSAEAGIFVLIYKMYGSEMLHK 225

RESULT 11

AAE13283

ID AAE13283 standard; protein; 491 AA.

XX

AC AAE13283;

XX

DT 12-FEB-2002 (first entry)
 XX
 DE Human transporters and ion channels (TRICH)-10.
 XX
 KW Human; transporter and ion channel; TRICH; akinesia; cystic fibrosis;
 KW diabetes mellitus; Parkinson's disease; myasthenia gravis; dementia;
 KW cardiac disorder; angina; hypertension; myocarditis; hyperglycaemia;
 KW neurological disorder; Alzheimer's disease; cataract; infertility;
 KW Wilson's disease; schizophrenia; Grave's disease; addison's disease;
 KW Huntington's disease; multiple sclerosis; meningitis; hypotensive;
 KW cardiant; nootropic; neuroprotective; neuroleptic; ophthalmological;
 KW antithyroid; anticonvulsant; goitre; antiinflammatory.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 241. .261
 FT /label= Transmembrane_domain
 FT Domain 251. .439
 FT /note= "Sodium, acid and bile transporter domain"
 FT Domain 288. .307
 FT /label= Transmembrane_domain
 FT Domain 325. .343
 FT /label= Transmembrane_domain
 FT Domain 416. .435
 FT /label= Transmembrane_domain
 XX
 PN WO200177174-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-US011206.
 XX
 PR 06-APR-2000; 2000US-0195595P.
 PR 12-APR-2000; 2000US-0196872P.
 PR 20-APR-2000; 2000US-0199020P.
 PR 28-APR-2000; 2000US-0200552P.
 PR 05-MAY-2000; 2000US-0202348P.
 PR 11-MAY-2000; 2000US-0203495P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Reddy R, Thornton M, Borowsky ML, Tang YT, Khan FA, Tribouley CM;
 PI Gandhi AR, Yao MG, Sanjanwala MS, Baughn MR, Nguyen DB, Policky JL;
 PI Yue H, Seilhamer JJ, Walia NK, Lal P, Kearney L, Walsh RT, Lu DAM;
 PI Lu Y, Greene BD, Raumann BE, Patterson C;
 XX
 DR WPI; 2002-017448/02.
 DR N-PSDB; AAD22002.
 XX
 PT Polypeptides of human transporters and ion channels, useful for
 PT diagnosing, treating or preventing disorders of transport, neurological,
 PT muscle, immunological and cell proliferative disorders.
 XX
 PS Claim 1; Page 130-131; 150pp; English.
 XX
 CC The invention relates to human transporters and ion channels (TRICH) and

CC the polynucleotides encoding them. The composition comprising TRICH or
 CC agonist of TRICH is useful for treating a disease or condition associated
 CC with decreased expression of functional TRICH or condition associated
 CC with overexpression of TRICH respectively. The composition comprising Ab
 CC is useful for diagnosing a condition of disease associated with
 CC expression of TRICH in a subject, where the disorders include a transport
 CC disorder such as akinesia, cystic fibrosis, diabetes mellitus,
 CC Parkinson's disease, myasthenia gravis, cardiac disorders associated with
 CC transport e.g. angina, hypertension, myocarditis, neurological disorders
 CC associated with transport e.g. Alzheimer's disease, Wilson's disease,
 CC schizophrenia, cataracts, infertility, hyperglycaemia, Grave's disease,
 CC goitre, addison's disease, Huntington's disease, dementia, multiple
 CC sclerosis, bacterial and viral meningitis. TRICH DNA is useful for
 CC generating a transcript image of a tissue or cell type, which represents
 CC the global pattern of gene expression by a particular tissue or cell type
 CC and for analysing the proteome of a tissue or cell type. TRICH DNA is
 CC used in gene therapy. The present amino acid sequence is human TRICH10
 CC protein

XX

SQ Sequence 491 AA;

Query Match 19.5%; Score 386.5; DB 5; Length 491;
 Best Local Similarity 27.1%; Pred. No. 1.3e-32;
 Matches 95; Conservative 56; Mismatches 105; Indels 95; Gaps 7;

Qy 44 LLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQA--IAVL 101
 : | |||:|: :|:| | | : | | | : : : |
 Db 115 ITMLGLGCTVDVNHFGAHVRRP---VAALLAALPVRPPAAAGLPAGPRLQAGRGGRRGLL 171

Qy 102 IMGCCPGGTISNIFTFVWDGMDL----- 125
 : ||||| :||: : ||||:|
 Db 172 LCGCCPGGNLSNLSLLVDGDMNLRRAALLALSSDVGSAQTSTPGLAVSPFHLYSTYKKK 231

Qy 126 -----SISMTTCSTVAALGMMPLCIYLYTWSW---SLQQNLTIPIYQ 163
 || || ||: || :|||:|:|:| : | :|
 Db 232 VSWLFDSKLVLLISAHSLFCSIIMTISSTLLALVLMPLCLWIYSWAWINTPIVQ--LLPLG 289

Qy 164 NIGITLVCLTIPVAFGVVNYRWPKQSKIILKI-----GAVVGGVLLL 207
 : :|| ||: ||:| :|: : : |:| : :| ||
 Db 290 TVTLTLCSTLIPIGLGVFIKYRSRVADYIVKVSLWSLLVTLVVLFIIMTGTMLGPELLAS 349

Qy 208 VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQNI 267
 : | |:| || |:|:| || | ||: |||:|:|
 Db 350 IPAAYVIA-----IFMPLAGYASGYGLATLFLHPPNCKRTVCLETGSQNV 395

Qy 268 QMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKRRLLKNK 318
 |:| :|:|:| : : | || | || : : | |:| : :|
 Db 396 QLCTAILKLAFFPPQFIGSMYMFLLYALFQSAEAGIFVLIYKMYGSEMLHK 446

RESULT 12

ABP43962

ID ABP43962 standard; protein; 491 AA.

XX

AC ABP43962;

XX

DT 26-FEB-2003 (first entry)

XX
 DE clone IMAGE:3502817.
 XX
 KW Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;
 KW gene therapy; nutritional supplement; wound; burn; ulcer;
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW autoimmune disorder; inflammation; vulnerary.
 XX
 OS Homo sapiens.
 XX
 PN WO200231111-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 11-OCT-2001; 2001WO-US027760.
 XX
 PR 12-OCT-2000; 2000US-00687527.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 DR WPI; 2002-426278/45.
 DR N-PSDB; ABQ61206.
 XX
 PT New polypeptides and their encoded proteins, useful as nutritional
 PT sources or supplements, or in gene therapy, particularly for treating
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
 PT inflammation.
 XX
 PS Claim 20; SEQ ID # 865; 357pp + Sequence Listing; English.
 XX
 CC The invention relates to 446 newly isolated polynucleotide sequences. The
 CC activity of polynucleotides of the invention may be described as,
 CC vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
 CC inflammatory. Compositions comprising nucleic acids of the invention are
 CC useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records ABP43544-
 CC ABP43989 represent polypeptides encoded by polynucleotides of the
 CC invention. NOTE: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 491 AA;

Query Match 19.3%; Score 381; DB 5; Length 491;
 Best Local Similarity 26.6%; Pred. No. 5.2e-32;
 Matches 98; Conservative 57; Mismatches 113; Indels 100; Gaps 8;

Qy 27 HGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLL 86
 || | : |:| |||:|: :|:| | | : | | |
 Db 103 HGLNVFVGAALCITMLG-----LGCTVDVNHFGAHVRRP---VAALLAALPVRPPAAAGL 154

XX
 PN WO200286076-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002WO-US012636.
 XX
 PR 19-APR-2001; 2001US-0284962P.
 PR 26-APR-2001; 2001US-0286645P.
 PR 09-JAN-2002; 2002US-0346986P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Carman J, Feder J, Nadler S;
 XX
 DR WPI; 2003-093119/08.
 DR N-PSDB; ACA54634.
 XX
 PT Novel NF-kappaB-associated polypeptides and polynucleotides useful for
 PT diagnosing, treating and preventing cancer, hepatic disorders, aberrant
 PT apoptosis, viral infections, autoimmune disorders, asthma and stroke.
 XX
 PS Claim 6; Page 488-489; 608pp; English.
 XX
 CC The present invention relates to the isolation of human nuclear factor-
 CC kappaB (NF-kappaB) associated polypeptides and polynucleotides. The NF-
 CC kappaB associated polypeptide and polynucleotide sequences are useful for
 CC preventing, treating or ameliorating various disorders including immune
 CC disorders, inflammatory disorders, cancers, disorders relating to
 CC aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas,
 CC haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal
 CC dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, al
 CC incontinentia pigmenti, viral infections (e.g. those caused by human
 CC immunodeficiency virus (HIV), human T-cell lymphotropic virus (HTLV),
 CC hepatitis B, hepatitis C, Epstein Barr virus (EBV), influenza),
 CC rheumatoid arthritis, inflammatory bowel disease, colitis, asthma,
 CC atherosclerosis, cachexia, euthyroid sick syndrome, stroke, experimental
 CC allergic encephalomyelitis (EAE), autoimmune disorders, disorders related
 CC to hyper immune activity, disorders related to aberrant acute phase
 CC responses, hypercongenital conditions, birth defects, necrotic lesions,
 CC wounds, organ transplant rejection, disorders related to aberrant signal
 CC transduction, hyperproliferative disorders, diseases of the pancreas
 CC (e.g. diabetes mellitus, vitamin B12 malabsorption), neurological
 CC disorders (e.g. Huntington's chorea), Turner's syndrome, bacterial
 CC infections, cardiovascular disorders, infertility, psoriasis and
 CC haemolytic anaemia. The present sequence represents a human NF-kappaB
 CC associated polypeptide of the invention
 XX
 SQ Sequence 490 AA;

Query Match 19.2%; Score 380.5; DB 6; Length 490;
 Best Local Similarity 26.8%; Pred. No. 5.8e-32;
 Matches 94; Conservative 56; Mismatches 106; Indels 95; Gaps 7;

Qy 44 LLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQA--IAVL 101
 : | |||:|:: :|:| | | : | | | : : :|
 Db 114 ITMLGLGCTVDVNHFGAHVRRP---VAALLAALPVRPPAAAGLPAGPRLQAGRGRRGLL 170

Qy 102 IMGCCPGGTISNIFTFWVDGMDL----- 125
 : ||||| :||: : ||||:|
 Db 171 LCGCCPGGNLSNLSLLVDGMNLRRAALLALSSDVGSAQTSTPGLAVSPFHLYSTYKKK 230

 Qy 126 -----SISMTTCSTVAALGMMPLCIYLYTWSW---SLQQNLTIPIYQ 163
 || || ||: || :|||::|:|:| : | :|
 Db 231 VSWLFDSKLVLSAHSLECSIIMTISSTLLALVLMPLCLWIYSWAWINTPIVQ--LLPLG 288

 Qy 164 NIGITLVCLTIPVAFGVVNYRWPKQSKIILKI-----GAVVGGVLLLV 207
 : :|| ||: ||:: |:: : : |::| : :| ||
 Db 289 TVTLTLCSTLIPIGLVGFIRYKYSRVADYIVKVSLWSLLVTLVVLFIIMTGTMLGPELLAS 348

 Qy 208 VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQNI 267
 : | |:| || : :|| | ||: |||:|:
 Db 349 IPAAYVIA-----IFMPLAAYASGYGLATLFLHLPNCKRRTVCLETGSQNV 394

 Qy 268 QMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKRRLKNK 318
 |:| :|:|:| : : | || | || : : | |:| : :|
 Db 395 QLCTAILKLAFPPQFIGSMYMFPLLYALFQSAEAGIFVLIYKMYGSEMLHK 445

RESULT 14

ABU69621

ID ABU69621 standard; protein; 490 AA.

XX

AC ABU69621;

XX

DT 05-JUN-2003 (first entry)

XX

DE Human NF-kappaB associated polypeptide sequence #24.

XX

KW Human; nuclear factor-kappaB; NF-kappaB; immune disorder; cancer;
 KW inflammatory disorder; apoptosis; hepatic disorder; Hodgkin's lymphoma;
 KW haematopoietic tumour; hyper-IgM syndrome; viral infection; asthma;
 KW hypohidrotic ectodermal dysplasia; human immunodeficiency virus; HIV;
 KW X-linked anhidrotic ectodermal dysplasia; al incontinentia pigmenti;
 KW influenza; rheumatoid arthritis; inflammatory bowel disease; colitis;
 KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;
 KW experimental allergic encephalomyelitis; autoimmune disorder; wound;
 KW hyper immune activity; acute phase response; hypercongenital condition;
 KW birth defect; necrotic lesion; organ transplant rejection; pancreas;
 KW signal transduction; hyperproliferative disorder; diabetes mellitus;
 KW vitamin B12 malabsorption; neurological disorder; Huntington's chorea;
 KW Turner's syndrome; bacterial infection; cardiovascular disorder;
 KW infertility; psoriasis; haemolytic anaemia; antiinflammatory; anti-HIV;
 KW cytostatic; hepatotropic; virucide; antirheumatic; antiarthritic;
 KW antiasthmatic; immunomodulator; antidiabetic; antiallergic;
 KW neuroprotective; immunosuppressive; vulnerary; antibacterial;
 KW antiinfertility; antianaemic; antipsoriatic; cerebroprotective; cardiant;
 KW antiarteriosclerotic.

XX

OS Homo sapiens.

XX

PN WO200286076-A2.

XX

PD 31-OCT-2002.

XX
PF 19-APR-2002; 2002WO-US012636.
XX
PR 19-APR-2001; 2001US-0284962P.
PR 26-APR-2001; 2001US-0286645P.
PR 09-JAN-2002; 2002US-0346986P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Carman J, Feder J, Nadler S;
XX
DR WPI; 2003-093119/08.
DR N-PSDB; ACA54634.
XX
PT Novel NF-kappaB-associated polypeptides and polynucleotides useful for
PT diagnosing, treating and preventing cancer, hepatic disorders, aberrant
PT apoptosis, viral infections, autoimmune disorders, asthma and stroke.
XX
PS Claim 4; Page 489-490; 608pp; English.
XX
CC The present invention relates to the isolation of human nuclear factor-
CC kappaB (NF-kappaB) associated polypeptides and polynucleotides. The NF-
CC kappaB associated polypeptide and polynucleotide sequences are useful for
CC preventing, treating or ameliorating various disorders including immune
CC disorders, inflammatory disorders, cancers, disorders relating to
CC aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas,
CC haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal
CC dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, al
CC incontinentia pigmenti, viral infections (e.g. those caused by human
CC immunodeficiency virus (HIV), human T-cell lymphotropic virus (HTLV),
CC hepatitis B, hepatitis C, Epstein Barr virus (EBV), influenza),
CC rheumatoid arthritis, inflammatory bowel disease, colitis, asthma,
CC atherosclerosis, cachexia, euthyroid sick syndrome, stroke, experimental
CC allergic encephalomyelitis (EAE), autoimmune disorders, disorders related
CC to hyper immune activity, disorders related to aberrant acute phase
CC responses, hypercongenital conditions, birth defects, necrotic lesions,
CC wounds, organ transplant rejection, disorders related to aberrant signal
CC transduction, hyperproliferative disorders, diseases of the pancreas
CC (e.g. diabetes mellitus, vitamin B12 malabsorption), neurological
CC disorders (e.g. Huntington's chorea), Turner's syndrome, bacterial
CC infections, cardiovascular disorders, infertility, psoriasis and
CC haemolytic anaemia. The present sequence represents a human NF-kappaB
CC associated polypeptide of the invention
XX
SQ Sequence 490 AA;

Query Match 19.2%; Score 380.5; DB 6; Length 490;
Best Local Similarity 26.8%; Pred. No. 5.8e-32;
Matches 94; Conservative 56; Mismatches 106; Indels 95; Gaps 7;

Qy 44 LLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQA--IAVL 101
: | |||:|:: :|:| | | : | | | : : : |
Db 114 ITMLGLGCTVDVNHFGAHRVRP---VAALLAALPVRPPAAAGLPAGPRLQAGRGRRGLL 170
Qy 102 IMGCCPGGTISNIFTFWVDGMDL----- 125
: ||||| :|: : ||||:|
Db 171 LCGCCPGGNLSNLSLLVDGDMNLRRAALLALSSDVGSAQTSTPGLAVSPFHLYSTYKKK 230

```

Qy      126 -----SISMTTCSTVAALGMMPLCIYLYTWSW---SIQQNLTIPIYQ 163
           || ||  ||: || :|||::|:|:|  : |  :|
Db      231 VSWLFDSKLVLI SAHSLFCSIIMTISSTLLALVLMPLCLWIYSWAWINTPIVQ--LLPLG 288

Qy      164 NIGITLVCLTIPVAFGVYVNYRWPKQSKIILKI-----GAVVGGVLLLV 207
           : :||  ||: ||: |:: : :  |::  | :|  ||
Db      289 TVTLTLCSTLIPIGLGVFI RYKYSRVADYIVKVSLWSLLVTLVVL FIMTGTMLGPELLAS 348

Qy      208 VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL LALFTHQSWQRCRTISLETGAQNI 267
           : |  |:|  ||  : :|: ||  |  ||: |||:|:|
Db      349 IPAAVYVIA-----IFMPLAAYASGYGLATLFHLPNCKR TVCLETGSQNV 394

Qy      268 QMCITMLQLSFTAELVQMLSFPPLAYGLFQLIDGFLIVAAYQTYKRRLKNK 318
           |:|  :|:|:|  : :  |  ||| | |||  : :  |  |:|  : :|
Db      395 QLCTAILKLAFFPPQFIGSMYMFPLLYALFQSAEAGIFVLIYKMYGSEMLHK 445

```

RESULT 15

AAE21252

ID AAE21252 standard; protein; 225 AA.

XX

AC AAE21252;

XX

DT 01-JUL-2002 (first entry)

XX

DE Human gene 8 encoded secreted protein fragment, SEQ ID NO:117.

XX

KW Human; secreted protein; immune disorder; antiallergic; antirheumatic;
KW rheumatoid arthritis; breast neoplasia; breast cancer; antiarthritic;
KW neurological disease; Alzheimer's disease; Parkinson's disease; trauma;
KW Tourette syndrome; encephalitis; cytostatic; haemostatic; anaemia; mania;
KW antiinflammatory; ophthalmological; dermatological; immunostimulatory;
KW immunomodulatory; immunosuppressive; antibacterial; antipsoriatic;
KW gene therapy; autoimmune disease; Huntington's disease; meningitis;
KW demyelinating disease; peripheral neuropathy; congenital malformation;
KW spinal cord injury; peripheral neuropathy; ischaemia; perception;
KW multiple sclerosis; infarction; haemorrhage; schizophrenia; dementia;
KW depression; panic disorder; learning disability; ALS; feeding disorder;
KW hyperproliferative disorder; sleep pattern; cardiovascular disorder;
KW reproductive disorder; digestive system disorder; behavioural disorder.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 200

FT /label= Unknown

FT /note= "Xaa equals any of the naturally occurring L-amino acids"

FT Misc-difference 204

FT /label= Unknown

FT /note= "Xaa equals any of the naturally occurring L-amino acids"

FT Misc-difference 206

FT /label= Unknown

FT /note= "Xaa equals any of the naturally occurring L-amino acids"

FT

FT Misc-difference 210
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring L-amino
 FT acids"
 FT Misc-difference 214
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring L-amino
 FT acids"
 FT Misc-difference 217
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring L-amino
 FT acids"
 FT Misc-difference 218
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring L-amino
 FT acids"
 FT Misc-difference 222
 FT /label= Unknown
 FT /note= "Xaa equals any of the naturally occurring L-amino
 FT acids"
 XX
 PN WO200216390-A1.
 XX
 PD 28-FEB-2002.
 XX
 PF 17-JAN-2001; 2001WO-US001435.
 XX
 PR 18-AUG-2000; 2000US-0226282P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR;
 PI Olsen HS, Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH;
 PI Fiscella M, Ni J;
 XX
 DR WPI; 2002-304113/34.
 XX
 PT An isolated nucleic acid molecule (I) comprising a polynucleotide which
 PT encodes a polypeptide useful in the diagnosis and treatment of disorders
 PT e.g. immune disorders.
 XX
 PS Disclosure; Page 26; 504pp; English.
 XX
 CC AAD33692-AAD33736 represent cDNAs corresponding to 21 human secreted
 CC protein genes, and AAE21191-AAE21235 represent the proteins they encode.
 CC AAE21236-AAE21280 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 21 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of immune
 CC or autoimmune diseases e.g. AIDS (acquired immune deficiency syndrome),
 CC asthma, anaemia and rheumatoid arthritis, breast neoplasia and breast
 CC cancer, neurological diseases e.g. Alzheimer's disease, Parkinson's
 CC disease, Huntington's disease, Tourette syndrome, meningitis,

CC demyelinating disease, peripheral neuropathies, neoplasia, trauma,
CC congenital malformations, spinal cord injuries, toxic neuropathies
CC induced by neurotoxins, peripheral neuropathies, multiple sclerosis,
CC ischaemia and infarction, haemorrhages, schizophrenia, mania, dementia,
CC depression, panic disorder, learning disabilities, ALS, altered
CC behaviours e.g. disorders in feeding, sleep patterns, balance and
CC perception, encephalitis, disorders in cardiovascular, neural/ sensory,
CC reproductive and digestive systems, behavioural disorders and
CC hyperproliferative disorder. The present sequence represents human
CC secreted protein fragment referred to in the disclosure of the invention
XX
SQ Sequence 225 AA;

```

Qy      86 LAISFLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMDL SISMTTCSTVAALGMMPLCI 145
      ||::| | | }::||: ||||| :||: : ||||:| | | |::| | :|||:
Db      1 LALAFKLDEVA AVALLCGCCPGGNLSNLM SLLVDGDMNLSIIMTISSTLLALVLMPLCL 60

Qy     146 YLYTWSW---SLQQNLTI PYQNIGITLVCLTIPVAFGVVYNRWPKQSKIILKI----- 196
      ::|:|:| : | :| : :|| ||: ||:: |: : : |:|:
Db      61 WIYSWAWINTPIVQ--LLPLGTVTLTLCSTLIPIGLVFI RYKYSRVADYIVKVSLSWLL 118

Qy     197 -----GAVVGGVLLLLVVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLLAL 246
      | ::| || : | |:| || |: :|: ||
Db     119 VTLVVLFI MTGTMLGP ELLASIPAAVYVIA-----IFMPLAGYASGYGLAT 164

Qy     247 FTHQSWQRCRTISLETGAQNIQMCITMLQLSF 278
      | ||: ||||:| |:| :|:|:|
Db     165 LFHLPPNCKRTVCLETGSONVOLCTAILKLA F 196

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:34:43 ; Search time 23 Seconds
(without alignments)
846.217 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEL.....PGPMDCHRALEFPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	884	44.7	348	1	US-08-176-126B-2	Sequence 2, Appli
2	884	44.7	348	2	US-08-669-435-2	Sequence 2, Appli
3	884	44.7	348	5	PCT-US94-14431A-2	Sequence 2, Appli
4	860.5	43.5	348	1	US-08-176-126B-4	Sequence 4, Appli
5	860.5	43.5	348	2	US-08-669-435-4	Sequence 4, Appli
6	860.5	43.5	348	5	PCT-US94-14431A-4	Sequence 4, Appli
7	280.5	14.2	325	4	US-09-328-352-5100	Sequence 5100, Ap
8	271.5	13.7	315	4	US-09-252-991A-17715	Sequence 17715, A
9	267.5	13.5	323	4	US-09-540-236-2883	Sequence 2883, Ap
10	257.5	13.0	308	4	US-09-252-991A-23958	Sequence 23958, A
11	237.5	12.0	349	4	US-09-489-039A-8584	Sequence 8584, Ap

12	122	6.2	659	4	US-09-543-681A-6013	Sequence 6013, Ap
13	115	5.8	722	4	US-09-489-039A-8942	Sequence 8942, Ap
14	104	5.3	324	4	US-09-328-352-4577	Sequence 4577, Ap
15	103	5.2	697	4	US-09-976-594-489	Sequence 489, App
16	102.5	5.2	478	4	US-09-107-532A-7010	Sequence 7010, Ap
17	101	5.1	370	4	US-08-405-271A-21	Sequence 21, Appl
18	101	5.1	534	4	US-09-252-991A-27300	Sequence 27300, A
19	100.5	5.1	381	4	US-09-721-870-28	Sequence 28, Appl
20	100.5	5.1	424	4	US-09-134-001C-5009	Sequence 5009, Ap
21	99.5	5.0	364	4	US-09-252-991A-31716	Sequence 31716, A
22	99.5	5.0	385	4	US-08-981-700A-4	Sequence 4, Appli
23	99.5	5.0	387	3	US-08-993-088A-7	Sequence 7, Appli
24	99.5	5.0	387	4	US-08-993-424B-7	Sequence 7, Appli
25	99.5	5.0	387	4	US-09-595-549-10	Sequence 10, Appl
26	99.5	5.0	387	4	US-09-603-680-7	Sequence 7, Appli
27	99.5	5.0	387	4	US-08-899-112B-28	Sequence 28, Appl
28	99	5.0	362	4	US-09-252-991A-27993	Sequence 27993, A
29	99	5.0	444	4	US-09-543-681A-5355	Sequence 5355, Ap
30	98.5	5.0	344	3	US-09-110-116-4	Sequence 4, Appli
31	98	5.0	399	4	US-09-489-039A-9414	Sequence 9414, Ap
32	98	5.0	429	4	US-09-489-039A-13214	Sequence 13214, A
33	98	5.0	835	4	US-09-284-819-6	Sequence 6, Appli
34	98	5.0	835	4	US-09-262-537-12	Sequence 12, Appl
35	97	4.9	462	4	US-09-328-352-6888	Sequence 6888, Ap
36	96.5	4.9	464	4	US-09-543-681A-5655	Sequence 5655, Ap
37	96.5	4.9	623	4	US-09-107-532A-4726	Sequence 4726, Ap
38	96.5	4.9	3177	2	US-08-477-451-4	Sequence 4, Appli
39	96	4.9	323	4	US-09-489-039A-7408	Sequence 7408, Ap
40	96	4.9	401	4	US-09-543-681A-5305	Sequence 5305, Ap
41	96	4.9	521	3	US-08-956-322-4	Sequence 4, Appli
42	96	4.9	572	4	US-08-937-067-13	Sequence 13, Appl
43	96	4.9	594	4	US-09-489-039A-10622	Sequence 10622, A
44	96	4.9	652	3	US-09-110-116-1	Sequence 1, Appli
45	96	4.9	652	3	US-08-956-322-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-176-126B-2

; Sequence 2, Application US/08176126B

; Patent No. 5589358

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

```

;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/176,126B
;   FILING DATE:  29-DEC-1993
;   CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Parker, David L.
;   REGISTRATION NUMBER:  32,165
;   REFERENCE/DOCKET NUMBER:  WAKE:002/PAR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (512) 418-3000
;   TELEFAX:  (512) 474-7577
;   TELEX:  na
;   INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  348 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-176-126B-2

```

```

Query Match          44.7%;  Score 884;  DB 1;  Length 348;
Best Local Similarity  46.9%;  Pred. No. 3.5e-82;
Matches 164;  Conservative  74;  Mismatches 102;  Indels  10;  Gaps  4;

```

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | : : | : | | :| : | |:: |::||:|:|:| | | :
Db      3 NSSICNP NATICEGDSCIAPESNFNAILSVVMSTVLTILLALVMFSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| | | ||||:| | |::|::| : ||||: | | ||||| | | | :||| |
Db     63 RRPWGI VVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCCPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
      |||:||||| |: ||||| |:| | | : ||| :|| :|| | ||| :|:|
Db    123 MDLSVSM TTCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPVSIGMYV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLL VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:|:|:|:| ||||| |: | :|:|:| | :| : | | | |:| :| :|
Db    183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGF 302
      | | | | | :| | | | | | | | | | :| | | :| | | :| | :| :|
Db    243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302

Qy    303 LIVAAAYQTYKRRLKNKHGKKN SGCTEVCHTRKS--TSSRETNAFLEVNEE 350
      ::| | | : | | | | : | :| | :| :| :| :| :| :|
Db    303 ILLGAYVAYKK---CHGKNNT ELQEKTDNEMEPRSSFQETNKGFPDEK 348

```

RESULT 2

US-08-669-435-2

```

; Sequence 2, Application US/08669435
; Patent No. 5869265
; GENERAL INFORMATION:

```



```

; APPLICANT: Dawson, Paul A.
; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,435
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,126
; FILING DATE: 29-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE:002/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-669-435-2

```

```

Query Match          44.7%; Score 884; DB 2; Length 348;
Best Local Similarity 46.9%; Pred. No. 3.5e-82;
Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

```

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
        :|| | |:: | : : | : | | : | | ::: |::|::|::|:: | | :
Db      3 NSSICPNATICEGDSIAPESNFNAILSVMSTVLTILLALVMFSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
        ||||| || |||||::|| | ::|::| : ||||: || | ||||| || | :|||
Db     63 RRPWGVVVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCCPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
        |||:|||||:: |||||:::| | : ||| :|| :|| | ||| :|:|
Db    123 MDLSVSMTCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPVSIGMYV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLL VAVAGVVLAKGSWNSDITLLTISFIFPLIGHTGF 242

```

```

Db      183 NHKWPQKAKIILKIGSIAGAILIVLIAVVGILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242
Qy      243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGF 302
        ||      | | ||||:|||| | | | :| |||: | | : :||| | :||:
Db      243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302
Qy      303 LIVAAQTYKRRLKNKHGKKNSGCTEVCHTRKS--TSSRETNAFLEVNEE 350
        ::: || ||:      ||| |:      |      :| :|||      : :|:
Db      303 ILLGAYVAYKK----CHGKNNTLQEKTDNEMEPRSSFQETNKGFPDEK 348

```

RESULT 3

PCT-US94-14431A-2

; Sequence 2, Application PC/TUS9414431A

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version

; SOFTWARE: #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/14431A

; FILING DATE: 29-DEC-1994

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: USSN 08/176,126

; FILING DATE: 29-DEC-1993

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: PARKER, DAVID L.

; REGISTRATION NUMBER: 32,165

; REFERENCE/DOCKET NUMBER: WAKE005P--

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (713) 789-2679

; TELEX: 79-0924(1) GENERAL INFORMATION:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US94-14431A-2

Query Match

44.7%; Score 884; DB 5; Length 348;

Best Local Similarity 46.9%; Pred. No. 3.5e-82;
Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

```
Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | : : | : | | : | |::: |:| |:| |:| | | : | :
Db      3 NSSICNP NATICEGDS CIAPESNFNAILSVVMSTVLTILLALVMFSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| | | ||||:| | | :|:::| : | |||: | | | ||||| | | : ||||
Db     63 RRPWGIVVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCCPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
      |||:||||| |: |||||:::| | | : ||| :| | :| | | ||| :| |
Db    123 MDLSVSMTTCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPVSIGMYV 182

Qy    183 NYRWPQKSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:| |:|:::| ||||| |: | :|:::| | | : | : | : | | | |:| :| : |
Db    183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGF 302
      | | | | | | | | | | | | | | | : | | | : | | | | | : | | : | | : | |
Db    243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302

Qy    303 LIVAAAYQTYKRRLKNKHGKKNSGCTEVCHTRKS--TSSRETN AFLEVNEE 350
      ::| | | | : | | | | : | | | : | | : | | : | | : | |
Db    303 ILLGAYVAYKK----CHGKNNTLQEKTDNEMEPRSSFQETNKGFPDEK 348
```

RESULT 4

US-08-176-126B-4

; Sequence 4, Application US/08176126B

; Patent No. 5589358

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/176,126B

; FILING DATE: 29-DEC-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, David L.

; REGISTRATION NUMBER: 32,165

; REFERENCE/DOCKET NUMBER: WAKE:002/PAR

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-176-126B-4

```

```

Query Match          43.5%; Score 860.5; DB 1; Length 348;
Best Local Similarity 45.6%; Pred. No. 8.8e-80;
Matches 160; Conservative 68; Mismatches 104; Indels 19; Gaps 4;

```

```

Qy      5 CSSSSACPANSSEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR 64
      || :|      | :      | :| : | |:: |::|::|::|::| |::|
Db      14 CSGASCVPESNFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR 64

Qy      65 PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMD 124
      |||| || |||||::| | ::|::| : |::| |::|::|::| | | :|||
Db      65 PWGICVGFCLCQFGIMPLTGFI LSVAFDILPLQAVVVLIIGCCPGGTASNILAYWVDGDMD 124

Qy      125 LSISMTTCSTVAALGMMPLCIYLTWSWSLQONLTIPYQNIGITLVCLTIPVAFGVVNY 184
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      125 LSVSMTTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPVVSIGMFVNH 184

Qy      185 RWPQKSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLL 244
      :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      185 KWPQKAKIILKIGSIAGAILIVLIAVVGGLYQSAWIIAPKLWIIIGTIFPVAGYSLGFLL 244

Qy      245 ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGFLI 304
      |      | ||||:: || || |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      245 ARIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPEELNVVFTFPLIYSIFQLAFAAIF 304

Qy      305 VAAQTYKRRRLKNKHGKKNNGCTEVCHTRKSTSSRETNAFLEVNEEGAITP 355
      : | ||:      ||| : |      | : :::|::| | | |
Db      305 LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFQP 345

```

RESULT 5

US-08-669-435-4

```

; Sequence 4, Application US/08669435
; Patent No. 5869265

```

GENERAL INFORMATION:

```

; APPLICANT: Dawson, Paul A.
; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: US
; ZIP: 77210

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,435
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,126
; FILING DATE: 29-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE:002/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-669-435-4

```

```

Query Match          43.5%; Score 860.5; DB 2; Length 348;
Best Local Similarity 45.6%; Pred. No. 8.8e-80;
Matches 160; Conservative 68; Mismatches 104; Indels 19; Gaps 4;

```

```

Qy      5  CSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR 64
      || :|      |:      | :| : | |:: |::|::|::|::| |::|
Db      14  CSGASCVPESNFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR 64

Qy      65  PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDM 124
      |||| || ||||::|| | :|:::| : |::|: ||::|::|::| || | :||::|
Db      65  PWGICVGFLLCQFGIMPLTGFILSVAFDILPLQAVVLLIIGCCPGGTASNILAYWVDGDM 124

Qy      125  LSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNI GITLVCLTIPVAFGVVNY 184
      ||:|||||||: |||||::|| |      :: ||| ||| :|| | :||: |:|:|:
Db      125  LSVSMTTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPV SIGMEFVNH 184

Qy      185  RWPQKSKIILKIGAVVGGVLLLVA VAVGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLL 244
      :||:::||||||:: | :|:::| | :| : |      | | ||| :| : |||
Db      185  KWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWIIAPKLWIIGTIFPVAGYSLGFLL 244

Qy      245  ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLS FPLAYGLFQLIDGFLI 304
      |      | ||||:: ||| || |:| |::|::| | | : :|| | :||| :
Db      245  ARIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPEELNVVFTFPLIYSIFQLAFAAIF 304

Qy      305  VAA YQTYKRRLKNKHGKKN SGCTEVCHTRKSTSSRETNAFLEVNEEGAITP 355
      : | ||:      ||| : |      | : : :| : | | |
Db      305  LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFQP 345

```

RESULT 6
 PCT-US94-14431A-4
 ; Sequence 4, Application PC/TUS9414431A
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: United States of America
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 ; SOFTWARE: PatentIn Release #1.0, Version
 ; SOFTWARE: #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/14431A
 ; FILING DATE: 29-DEC-1994
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: USSN 08/176,126
 ; FILING DATE: 29-DEC-1993
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PARKER, DAVID L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: WAKE005P--
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (713) 789-2679
 ; TELEX: 79-0924(1) GENERAL INFORMATION:
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 348 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 PCT-US94-14431A-4

Query Match 43.5%; Score 860.5; DB 5; Length 348;
 Best Local Similarity 45.6%; Pred. No. 8.8e-80;
 Matches 160; Conservative 68; Mismatches 104; Indels 19; Gaps 4;

Qy	5	CSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR	64
		: : : : :: : : : : :	
Db	14	CSGASCVPESNFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR	64
Qy	65	PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDM	124
		: : : : : : : : : :	
Db	65	PWGICVGLFCQFGIMPLTGFILSVAFDILPLQAVVLIIGCCPGGTASNILAYWVDGDM	124

Qy	125	LSISM TTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNI GITLVLCTIPVAFGVVNY	184
Db	125	LSVSM TTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPV SIGMFVNH	184
Qy	185	RWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL	244
Db	185	KWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWIIAPKLWII GTIFPVAGYSLGFL	244
Qy	245	ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLS FPLAYGLFQLIDGFLI	304
Db	245	ARIAGLPWYRCRTVAFETGMONTQLCSTIVQLSFTPEELNVVFTFPLIYSIFQLAFAAIF	304
Qy	305	VAAYQTYKRRLKNKHGKKN SGCTEVCHTRKSTSSRETNAFLEVNEEGAITP	355
Db	305	LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFOP	345

US-09-328-352-5100

; Sequence 5100, Application US/09328352

; GENERAL INFORMATION:

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT FILING DATE: 1999-06-04

; SEQ ID NO 5100

```
; TYPE: PRT
```

US-09-328-352-5100

Best Local Similarity 25.7%; Pred. No. 1.9e-20;

Matches 71; Conservative 70; Mismatches 116; Indels 19; Gaps 7;

QY 41 MMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSCLKPVQAIAV 100
 ::||::| :| :: : :: | : : ||:: || :|| |::| :| | :| :

QY 101 LIMGCCPGGTISNIFTFWVDGMDLSISMTCSTVAALGMMPLCIYLTSWSLQQNLTI 160
 :::||| |||| | : : | ::| :|| :: | : |: | : | || |

Qy 161 PYQNIGITLV-CLTIPVAFGVYNYRWPKQSKIILKIGAVGGVLLLVAVAGVVLA KGS 219
 :: |::: : :: | : : : : | : : : : | : : : |

```
QY      220 ----WNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQNIQMCITMLQ 275
          | : : | : : : : : | | | : : | : | | : : :
```

Qy 276 LSFTA EHLVQMLS FPLAYGLFQ LIDGFLIVAAYQTY 311

: | | : : | : | : | | | |
Db 288 VHFAASPITAVPS--AIFSLWHNISG----PALATY 317

RESULT 8

US-09-252-991A-17715
; Sequence 17715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17715
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17715

Query Match 13.7%; Score 271.5; DB 4; Length 315;
Best Local Similarity 25.7%; Pred. No. 1.6e-19;
Matches 75; Conservative 74; Mismatches 116; Indels 27; Gaps 8;

Qy 32 LVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFS 91
 | | ::||:| | :| :: | | : :|:| || :|| |:||
Db 38 LPLTAAIAPLLGLVMFGMGLTLKGEDFREVARHPPIRVLIGVLAQFVIMPGWLWLLCSLLQ 97

Qy 92 LKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWS 151
 | |: |::| ||||| ||: |: ||: ||::|: |:| | :| :| :
Db 98 LPAEIAVGVLVGGCCPGGTASNVTWLSRGDVALSVAITSVTTLLAPLVT PALVWLLASA 157

Qy 152 WSLQQNLTI PYQNIGITLV-CLTIPVAFGVVNYRWPKQSKIILKIGAVVGGVLLLVAV 210
 | | : : : : : : :|:| | : ::: : : : :| | :|| :
Db 158 W-----LPVSFAAMFLSILQVVLVPIALGLLAQRLLGERTRQVAEVLPLV-SVFSIVVII 211

Qy 211 AGVVLAKGSWNSDITLLTISFIF-----PLIGHVTGFLALFTHQSWQRCRTISLETGA 264
 | || | : : : || : : : |:|::|| | : | : : :| |
Db 212 AAVVAASQARIAESGLLIMAVVMLHNGFGLLLGYLTGKLTGMPLAQR----KALAIEVGM 267

Qy 265 QNIQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAYQTYKRRL 315
 || : | : | : : | | : : : | | : | : |||
Db 268 QN-----SGLGAALANVHFAPLA AVPSALFSVWHNLSGSLLAALF----RRL 310

RESULT 9

US-09-540-236-2883
; Sequence 2883, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:

; LENGTH: 308
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23958

Query Match 13.0%; Score 257.5; DB 4; Length 308;
 Best Local Similarity 26.2%; Pred. No. 4.1e-18;
 Matches 75; Conservative 60; Mismatches 128; Indels 23; Gaps 7;

Qy 33 VFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSL 92
 : |: : :|::| || |: :| | : ||:| | |:| :|: |:|
 Db 17 ILTLFLPIALGIIMLGLSLTPADFLRVVRYPKPVLVGLVCQIVLLPLACFLIVQGFAL 76

Qy 93 KPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSW 152
 : |: :::: |||| :|::: ||: |:|:| :::| |: ||| :|
 Db 77 EAALAVGMMLLAASPGGTTANLYSHLAHGDVALNITLTAVNSVIAILTMPLIVNL----- 131

Qy 153 SLQ-----QNLTI PYQNIGITLVCLTIPVAFGVVNYRWPKQSKIILKIGAVVGGVLLL 206
 ||| | : : : : | : ||| |:| | | : : | : : :|||
 Db 132 SLQYFMGDGQAIPLQFGKVVQVFVIVLGPVAIGMLVRNRLPAVADRLQKPKVILSALLLL 191

Qy 207 VVAVAGVVLAKGSWNSDITLLTI----SFIFPLIGHVTGFLALFTHQSWQRCRTISLET 262
 |: : : ||| | : :| : : | |: |: : : : :| :|
 Db 192 VIIL--LALAK-DWQTFV TYAPVVGLAALAFNLLSLAVGYWVPRLLRLPKAQA VAIGMEI 248

Qy 263 GAQNIQMCITMLQLSFTA EHLVQMLS FPLAYGLFQLIDGFLIVAAY 308
 | | : | | || : : | | ||: | ||:
 Db 249 GIHNGTLAIA-LALSPSLLNNSTMAIPPAIYGVLM----FFTAAAF 289

RESULT 11

US-09-489-039A-8584
 ; Sequence 8584, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8584
 ; LENGTH: 349
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-8584

Query Match 12.0%; Score 237.5; DB 4; Length 349;
 Best Local Similarity 26.5%; Pred. No. 5.4e-16;
 Matches 79; Conservative 57; Mismatches 109; Indels 53; Gaps 9;

Qy 37 VSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQ 96
 |:|::| |:|| :| ::: :|| :| |: : :|| |:|::| :| :

Db	74	VTLLM-LIMFGMGVHLKLEDFKRVLSRPAPVAAGIFLHYLVMPLAAWLLALLFHMPEL	132
Qy	97	AIAVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQ	156
		: :::: : : ::: : :	
Db	133	SAGMVLVGSVASGTASNMI FLAKGDVALSVTISSVSTLVGVVATPLLTRLV-----	185
Qy	157	NLTIPYQNI GITLVCL---TIPVAFGVVNYRWPQSK-----IILKIGAVV	200
		: : : : : : :	
Db	186	DAHIQVDVMGMLLSILQIVVIPIALGLIVHLLPKVVKAVEPFLPAFSMVCILAIISAVV	245
Qy	201	GGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISL	260
		: : : : : : :::	
Db	246	AGSAAHIASVGLVVI IAVILHNTIGLL-----GGYWGGRLFGFDEST---CRTLAI	293
Qy	261	ETGAQNIQMCITMLQLSF TAEHLVQMLSFPLA-----YGLFQLIDGFLIVAAYQTYK	312
		: : : : : : : :	
Db	294	EVGMONSGLAALGKIYFG-----PLAALPGALFSVWHNLSGSL-AGYWSGK	340

Query Match 6.2%; Score 122; DB 4; Length 659;
Best Local Similarity 27.5%; Pred. No. 0.00089;
Matches 56; Conservative 24; Mismatches 66; Indels 58; Gaps 11;

Db :| : | |||: |: ||
193 ALTFKDTTITFGLIGY--AFISAL 214

RESULT 13

US-09-489-039A-8942
; Sequence 8942, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8942
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8942

Query Match 5.8%; Score 115; DB 4; Length 722;
Best Local Similarity 26.5%; Pred. No. 0.0053;
Matches 54; Conservative 26; Mismatches 66; Indels 58; Gaps 11;

```
Qy      67 GIAVG--LLCQFGLMPFTAYLLAISFSLKPVQAIAVLI-----MGCCP 107
          |  ||  |  |  | :|  | :||  |  |  |  |  |  |  |  |  |  |
Db      108 GPLVGPVLAAQMGYLPGLTLWLLAGVVLAVQDFMVLFISSRRNGASLGEMIKQEMGPVP 167

Qy      108 GGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNI 167
          |:|:  | :  | : ::  | ||  |  |  |  |  |  |  |  |  |
Db      168 -GSIALFGCFI---MIIILAVLALIVVKALAESP-----W-----GV 201

Qy      168 TLVCLTIPVA--FGVYVNYRWPQSKIILKIGAV-VGGVLLLIVAV--AGVVLAKGSWNS 222
          || |:|:|  |:|: :  |  :|  |  | |:|  |  |  |  |  |
Db      202 FTVCSTVPIALFMGIYMRFLRPG-----RVGEVSVIGIVLLVASIWFGGVIAHDPYWG 255

Qy      223 DITLLTISFIFPLIGHVTGFLAL 246
          :|       : | |||:   |: ||
Db      256 ALTFKDTTITFTLIGY--AFISAL 277
```

RESULT 14

US-09-328-352-4577
; Sequence 4577, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352

```
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4577
;   LENGTH: 324
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-4577
```

```

Query Match          5.3%; Score 104; DB 4; Length 324;
Best Local Similarity 23.6%; Pred. No. 0.023;
Matches 71; Conservative 48; Mismatches 124; Indels 58; Gaps 15;

Qy      26 VHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGI-----AVGL 72
      | |      | :|| : :| | | :      :      | | :      |:||
Db      29 VSGQAAQYFNTLTVAIAILFFLHGAKLSREAVIEGILH-WKMHLLVFAITFFIFPAIGL 87

Qy      73 LCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNI-FTFWVDGMDLSISMTT 131
      | : | :|      |      |      | |      | :|| |      | : : : :
Db      88 LAKPILLPLLGGQQLYWGF-----LFMCFLPSTVQSSIAFTSVAKGNVAGAVCSAS 137

Qy     132 CSTVAALGMMPLCIYLYTWSWSLQQNLTIPTYQNIGITLVCLTIPVAFGVYVN-YRWPKQS 190
      | : : : | : : :      | |      | ||| : | :|      | : | :| :
Db     138 FSNLVGMFITPVLVSFFILGQS-QHGFDPDTSSIIQITLL-LLVPFVLGQLLRPYVFPYMA 195

Qy     191 KIILKIGAVVGGVLLLVV--AVAGVVLAKGSWN----SDITLLTI--SFIFPLIGHVTGF 242
      |:      : |      | :||| | :| | :| | :      : |||| | : :|
Db     196 KVPSIVKAFDQGSILMVVYGAFSGAVVA-GLWHQVSWKTL LLLTIACSVLLTII-----M 249

Qy     243 LLALFTHQS--WQRCRTI-----SLETGAQNIQMC-----ITMLQLSFTAHLVQM 286
      |||| : : : |      |      | :| | : |      : | : |      | :| :
Db     250 LLALYLPRAFGENRADQITVFFCGSKKTLASGVPMAQILFIGQPLGMIVLPIMIFHQIQL 309

Qy     287 L 287
      :
Db     310 M 310

```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 2120743CD1
US-09-976-594-489

Query Match 5.2%; Score 103; DB 4; Length 697;
Best Local Similarity 22.0%; Pred. No. 0.085;
Matches 82; Conservative 58; Mismatches 141; Indels 92; Gaps 19;

```
Qy      22 VGLEVH-----GNLELVFTVVSTVMMGL-----LMFSL-GCSVEIRKLWSHIRRPWGIAV 70
      :|| |      :| ||  ::  ::||  ||:  |:  :  ||      :|  :
Db      50 LGLYVRWEKTANSLILVIFILGLFVLGSIILYYSFMEASLSLSNLW-----FGFLL 103

Qy      71 GLLCQFGLMPF-----TAYLLAISFSLKPVQAI AVLIMGCC---PGGTISNIFTFWV 119
      |||| |      | ||| |  |:  ::  | |      | |:  |
Db     104 GLLCFLDNSSEFKNDVKEESTKYLLLTISIVLRILCSLVERISGYVRHRP--TLLTTVEF-- 159

Qy     120 DGDMDL---SISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNI GITLVCLTI-- 174
      ::|  :|: ||  :| :: | : |  |:  :  |: |  | |
Db     160 ---LELVGFIASTTMLVEKSLSVILLVVALAMLIIDLRMKSFLAIPNLVIFAVLLFFSS 216

Qy     175 -----PVAE-----GVYVNYRWPK---QSKIILKIGAVVGGVLLLV 207
      |:||      |: | ||  : :| :: | |: |
Db     217 LETPKNPIAFACFFICLITDPFLDIYFSGLSVTERWKPFYRGRICRRLSVVFAGMIELT 276

Qy     208 VAVAGVVLAKGS--WNSDITLLTISFIFPLIGHVTGFLAL-----FTHQS 251
      :  : : | |  :| || :| |: ||| |  ||::
Db     277 FFILSAFKLRDTHLWYFVIPGFSIFGIFWMICHII-FLLTLWGFHTKLNDCHKVYFTHRT 335

Qy     252 -WQRCRTISLETGAQNIQMCITMLQL---SFTAHLVQMLSFPLAYGLFQLIDGFLIVAA 307
      :  |  | :: |: || | | :: :|:  |: | : |||
Db     336 DYNSLDRIMASKGMRH--FCLISEQLVFFSLLATAILGAVSWQPTNGIF--LSMFLIVLP 391

Qy     308 YQTYKRRLKNKHG 320
      ::  | :: |
Db     392 LESMAHGLFHELG 404
```

Search completed: March 23, 2004, 14:38:24
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:33:32 ; Search time 20 Seconds
(without alignments)
1813.210 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEEL.....PGPMDCHRALEPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	%		Query			Description
	Score	Match	Length	DB	ID	
1	884	44.7	348	2	A49876	Na+-dependent bile
2	860.5	43.5	348	2	I38655	ileal sodium-depen
3	559.5	28.3	362	2	A41601	Na+/taurocholate t
4	553	27.9	349	2	I55601	Na+/taurocholate co
5	333.5	16.9	477	2	S01696	gene P3 protein -
6	325	16.4	321	2	E69902	probable sodium-de
7	301.5	15.2	305	2	D90031	hypothetical prote
8	299.5	15.1	318	2	AD3295	sodium/bile acid c
9	299.5	15.1	323	2	B83757	sodium-dependent t
10	284	14.4	338	2	T02645	hypothetical prote
11	269.5	13.6	311	2	F83236	probable transport
12	266.5	13.5	315	2	B81168	transporter NMB070
13	265.5	13.4	315	2	E81937	probable transmemb

14	257.5	13.0	297	2	D83438	probable transport
15	250	12.6	297	2	E70482	Na(+) dependent tr
16	238.5	12.1	292	2	S75443	P3 protein - Synec
17	223	11.3	207	2	AH3120	sodium bile acid s
18	223	11.3	207	2	G98166	probable transport
19	222.5	11.2	318	2	G81995	probable transmemb
20	219.5	11.1	298	2	AF2462	hypothetical prote
21	180.5	9.1	182	2	I54222	housekeeping prote
22	171	8.6	460	2	T04579	hypothetical prote
23	156	7.9	379	2	T06610	hypothetical prote
24	142	7.2	344	2	T34981	probable integral
25	125	6.3	409	2	I40489	hypothetical prote
26	121.5	6.1	324	2	H72285	conserved hypothet
27	121	6.1	612	2	T40506	major facilitator
28	120	6.1	716	2	AB1070	probable carbon st
29	117	5.9	721	2	S56580	carbon starvation
30	117	5.9	721	2	A98293	probable carbon st
31	117	5.9	721	2	D86134	probable carbon st
32	116	5.9	286	2	JS0384	hypothetical 30.5K
33	115	5.8	559	2	A81752	conserved hypothet
34	112	5.7	349	2	AG3010	hypothetical prote
35	112	5.7	349	2	G98273	iron(III) dicitrat
36	112	5.7	656	2	A84018	hypothetical prote
37	110.5	5.6	467	2	AF0488	amino acid permeas
38	110.5	5.6	469	2	F97326	probable MDR-type
39	110	5.6	559	2	E71490	probable efflux pr
40	109.5	5.5	383	2	S76146	hypothetical prote
41	109	5.5	593	2	B90144	hypothetical prote
42	107.5	5.4	390	1	A69746	chloramphenicol re
43	107.5	5.4	576	2	T12005	NADH2 dehydrogenas
44	106.5	5.4	406	2	AF0097	probable membrane
45	106.5	5.4	621	2	E90253	formate hydrogenly

ALIGNMENTS

RESULT 1

A49876

Na+-dependent bile acid transporter, ileal - golden hamster

C;Species: *Mesocricetus auratus* (golden hamster)

C;Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 05-Nov-1999

C;Accession: A49876

R;Wong, M.H.; Oelkers, P.; Craddock, A.L.; Dawson, P.A.

J. Biol. Chem. 269, 1340-1347, 1994

A;Title: Expression cloning and characterization of the hamster ileal sodium-dependent bile acid transporter.

A;Reference number: A49876; MUID:94117449; PMID:8288599

A;Accession: A49876

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-348 <WON>

A;Cross-references: GB:U02028; NID:g455032; PIDN:AAA18640.1; PID:g455033

C;Keywords: transmembrane protein

Query Match	44.7%;	Score 884;	DB 2;	Length 348;
Best Local Similarity	46.9%;	Pred. No. 2.5e-65;		

Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

```
Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |: | : : | : | | : : | |:: |:| |:| |:| : | | :
Db      3 NSSICNPNATICEGDSCTAPESNFNAILSVVMSTVLTILLALVMFSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| || |||||:| | :|:|:| : | |||: || ||||| || | :||| |
Db     63 RRPWGI VVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCCPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
      |||:||||| |: |||||:|:| | | : || :| :| | ||| :|:|
Db    123 MDLSVSMTCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPVSIGMYV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVV LAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:| |:|:| ||||| |:| :|:|:| | :| : :| : | | |:| :| : ||
Db    183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGF 302
      || | | |||:| || | | |:| |:| |:| : | | : :|| | :|:|
Db    243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302

Qy    303 LIVAAAYQTYKRRLKNKHGKNSGCTEVCHTRKS---TSSRETN AFLEVNEE 350
      ::| | |: | | |: | | :| :|| : :| :
Db    303 ILLGAYVAYKK---CHGKNNTLQEKTDNEMEPRSSFQETNKGFPDEK 348
```

RESULT 2

I38655

ileal sodium-dependent bile acid transporter - human

C;Species: Homo sapiens (man)

C;Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000

C;Accession: I38655

R;Wong, M.H.; Oelkers, P.; Dawson, P.A.

J. Biol. Chem. 270, 27228-27234, 1995

A;Title: Identification of a mutation in the ileal sodium-dependent bile acid transporter gene that abolishes transport activity.

A;Reference number: I38655; MUID:96070831; PMID:7592981

A;Accession: I38655

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-348 <RES>

A;Cross-references: EMBL:U10417; NID:g2623285; PIDN:AAC51870.1; PID:g595399

A;Experimental source: Crohn's disease patient (heterozygous)

A;Note: the wild type is shown; a form with 290-Ser was deficient in transport activity

C;Genetics:

A;Gene: SLC15-A2

Query Match 43.5%; Score 860.5; DB 2; Length 348;

Best Local Similarity 45.6%; Pred. No. 2.2e-63;

Matches 160; Conservative 68; Mismatches 104; Indels 19; Gaps 4;

```
Qy      5 CSSSSACPANSSEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRR 64
      || :| | : | : | | : : | |:: |:| |:| |:| : | | :
Db     14 CSGASCVVPE SNFNFI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR 64
```

Qy 65 PWGIAVGLLCQFGIMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMD 124
 ||| || ||||:| | :|:|:| : | |:| : ||:||||| || :|||
 Db 65 PWGICVGFLLCQFGIMPLTGFI LSVAFDILPLQAVVLIIGCCPGGTASNILAYWVDGMD 124

Qy 125 LSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYVNY 184
 ||:|||||: |||||: :| | :| || | :| | :|:| :|:|:
 Db 125 LSVSMTTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPYDNIGTSLVALVVPV SIGMFVNH 184

Qy 185 RWPQSKIIILKIGAVVGGVLLL VVAVAGVV LAKGSWNSDITLLTISFIFPLIGHVTGFL 244
 :|:|:| |||||: | :|:|:| | :| : :| | | ||: | : |||
 Db 185 KWPQKAKIIILKIGSIAGAILIVLIAVVG GILYQSAWIIAPKLWIIGTIFPVAGYSLGFL 244

Qy 245 ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPLAYGLFQLIDGFLI 304
 | | |||:| || | | |:| |:|||| | | : :|| | :||| :
 Db 245 ARIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFT PEELNVVFTFPLIYSIFQLAFAAIF 304

Qy 305 VAA YQTYKRRLKNKHGKNSGCTEVCHTRKSTSSRETN AFLEVNEEGAITP 355
 : | ||: || : | : :| : | | |
 Db 305 LGFYVAYKK----CHGKNKAEIPE----SKENGTEPE SSFYKAN--GGFQP 345

RESULT 3

A41601

Na+/taurocholate transport protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Feb-1997

C;Accession: A41601

R;Hagenbuch, B.; Stieger, B.; Foguet, M.; Luebbert, H.; Meier, P.J.

Proc. Natl. Acad. Sci. U.S.A. 88, 10629-10633, 1991

A;Title: Functional expression cloning and characterization of the hepatocyte Na(+)/bile acid cotransport system.

A;Reference number: A41601; MUID:92073340; PMID:1961729

A;Accession: A41601

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-362 <HAG>

A;Cross-references: GB:M77429

C;Keywords: transmembrane protein

Query Match 28.3%; Score 559.5; DB 2; Length 362;

Best Local Similarity 37.2%; Pred. No. 1.4e-38;

Matches 133; Conservative 69; Mismatches 135; Indels 21; Gaps 9;

Qy 10 ACPANSSEELPVGLEVHGNLELVFTTVVSTVMMGLIMFSLGCSVEIRKLWSHIRRPWGIA 69
 : | | | || | | : :|: |:| |||:| | :|:| :| |:
 Db 7 SAPFNFS---LPPGFG-HRATDKALSIILVLM LLLIMLSLGCTMEFSKIKAH LWKPKGVI 62

Qy 70 VGLLCQFGIMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISM 129
 | |: |||:| | |:| | | :|:|:| | || | |:|:| : |||:| | |
 Db 63 VALVAQFGIMPLAAFL LGKIFHLSNIEALAILICGCSPGGNLSNLFTLAMKGDMNLSIVM 122

Qy 130 TTCSTVAALGMMPLCIYLYT---WSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYVNYRW 186
 ||||: :||||| :|:|: : | : :|:| | |:| : | | : : :
 Db 123 TTCSSFSALGMMPLLLYVYSKGIYDGD LDKD--VPYKGIMISLVIVLIPCTIGIVLKS KR 180

Qy 187 PKQSKIIILKIGAVVGGVLLL VVAVAGVV LAKGSWNSDIT--LLTISFIFPLIGHVTGFL 244
 | || | :| :| | : | :| || | :| | :|:|

```

Db      181 PHYVPYILKGGMIITFLLSVAVTALSVINVGNSIMFVMTPHLLATSSSLMPFSGFLMGYIL 240
Qy      245 -ALFTHQSWQRC-RTISLETGAQNIQMCITMLQLSFTAHLVQMLSFFPLAYGLFQLIDGF 302
      ||| | | |||:||| |||:| |:| ::| | : : ||| | :||| :|
Db      241 SALF--QLNPSCRRTISMETGFQNIQLCSTILNVTFPPEVIGPLFFFPLLYMIFQLAEG 298
Qy      303 LIVAAAYQTYKRRLKKNKHGKKNKSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
      ||: :: |:| | | |:: : :| | : | | | |||
Db      299 LIIIIIFRCYEKI-----KPPKDQTKITYKAAATEDATPAALEKGTHNGNIPPLQPGP 350

```

RESULT 4

I55601

Na/taurocholate cotransporting polypeptide - human

C;Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C;Accession: I55601

R;Hagenbuch, B.; Meier, P.J.

J. Clin. Invest. 93, 1326-1331, 1994

A;Title: Molecular cloning, chromosomal localization, and functional characterization of a human liver Na/bile acid cotransporter.

A;Reference number: I55601; MUID:94179485; PMID:8132774

A;Accession: I55601

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-349 <RES>

A;Cross-references: GB:L21893; NID:g410213; PIDN:AAA36381.1; PID:g410214

C;Genetics:

A;Gene: GDB:SLC10A1; NTCP

A;Cross-references: GDB:344932; OMIM:182396

A;Map position: 14pter-14qter

Query Match 27.9%; Score 553; DB 2; Length 349;

Best Local Similarity 36.0%; Pred. No. 4.6e-38;

Matches 124; Conservative 77; Mismatches 109; Indels 34; Gaps 10;

```

Qy      31 ELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAIF 90
      :| :|: | : :| |||:| | : :| :|: | : :|:| | :|
Db      24 DLALSIVLVFMLFFIMLSLGCTMEFSKIKAHLLWKPKGLAIALVAQYGIMPLTAFVLGKVF 83
Qy      91 SLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISM TTCSTVAALGMMPLCIYLYT- 149
      || :|:|:|: || ||| :|:|: : |||:| | ||| | ||| | :|:
Db      84 RLKNIEALAILVCGCSPGGNLSNVFSLAMKGD MNLSIVMTTCSTFCALGMMPLLLYIYSR 143
Qy      150 --WSWSLQQNLTI PYQNIGITLVCLTIPVAFGVVYNYRWPKQSKIILKIGAVVGGVLLLV 207
      : | : :|:| | | :| : || : : : | : : :| | : : :|
Db      144 GIYDGDLDKDK--VPYKGIVISLVVLIPCTIGIVLKS KRPQYMRVVIKGGMII----ILL 197
Qy      208 VAVAGVVLAKGSWN SDI-----TLLTISFIFPLIGHVTGFL-ALFTHQSWQRC-RTIS 259
      :|| ||: : | | : | : | | : |:| || | ||:|
Db      198 CSVAVTVLSAINVGK SIMFAMTPLL IATSSLM PFI GLLGYVLSALFCLNG--RCRRTVS 255
Qy      260 LETGAQNIQMCITMLQLSFTAHLVQMLSFFPLAYGLFQLIDGFLIVAAAYQTYKRRLKKNH 319
      :||| ||:|:| |:| ::| | : : ||| | :||| :| |:| : | : : |
Db      256 METGCQNVQLCSTILNVAFPPEVIGPLFFFPLLYMIFQLGEGLLLIAIFWCYE-KFKTPK 314
Qy      320 GKKNKSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGPMDC 363

```

Db 315 DK-----TKMIYTAATT-----EETIPGALNGTYKGEDC 344

RESULT 5

S01696

gene P3 protein - human

C;Species: Homo sapiens (man)

C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 05-Nov-1999

C;Accession: S01696

R;Alcalay, M.; Toniolo, D.

Nucleic Acids Res. 16, 9527-9543, 1988

A;Title: CpG islands of the X chromosome are gene associated.

A;Reference number: S01696; MUID:89041548; PMID:3186440

A;Accession: S01696

A;Molecule type: DNA

A;Residues: 1-477 <ALC>

A;Cross-references: EMBL:X12458; NID:g35187; PIDN:CAA30998.1; PID:g35188

Query Match 16.9%; Score 333.5; DB 2; Length 477;
Best Local Similarity 31.5%; Pred. No. 7.6e-20;
Matches 87; Conservative 53; Mismatches 115; Indels 21; Gaps 3;

Qy 12 PANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVG 71
|| : | | :: : : | || ||: | :: | : :|
Db 172 PAEDTPATLSADLAHFSENPIYLLPLIFVNKCSF--GCKVELEV LKGLMQSPQPMLLG 229
Qy 72 LLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTT 131
|| || :|| |:| | | |: :| || | :|: : ||: |:||||
Db 230 LLGQFLVMPLYAFIMAKVFMLPKALALGLIITCSSPGGGGSYLF SLLLGGDVT LAISMTF 289
Qy 132 CSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNGITLVCLTIPVAFGVYVNYRWPKQSK 191
|||| | :|| :|: |: : | :| | ||: : ||:| || : : || |:
Db 290 LSTVAATGFLPLSSAIYSRLLSIHETLHVPISKILGTLLFIAIPIAVGVLIKSKLPKFSQ 349
Qy 192 IILKIGAVVGGVLLL-----VVAVAGVV LAKGSWNSDITLLTISFIFPLIGHVTG 241
: :||: |||| | :||: | : : : ||:| : |
Db 350 LLLQVVKPFSEFVLLLGGFLAYRMGVFILA GIRL-----PIVLVGITVPLVGLLVG 400
Qy 242 FLLALFTHQSWQRCRTISLETGAQNIQMCITMLQLS 277
: || : ||:| | || : : |||||
Db 401 YCLATCLKLPVAQRRTVSIEVGVQNSLLALAMLQLS 436

RESULT 6

E69902

probable sodium-dependent transporter yocS - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: E69902

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: E69902

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-321 <KUN>

A;Cross-references: GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13827.1; PID:g2634328

A;Experimental source: strain 168

C;Genetics:

A;Gene: yocS

C;Superfamily: *Bacillus subtilis* sodium-dependent transporter yocS

Query Match 16.4%; Score 325; DB 2; Length 321;
Best Local Similarity 27.6%; Pred. No. 2.5e-19;
Matches 84; Conservative 76; Mismatches 114; Indels 30; Gaps 12;

Qy	33	VFTVVS---TVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAIS	89
		: : : :: : :: : :: : :: : :	
Db	32	LFTWISSYITIFLGIIMFGMGLTLQADDFKELVRKPWQVIIGVIAQYTIMPLVAFGLAFG	91
Qy	90	FSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYT	149
		: :: : : :: : :	
Db	92	LHLP AEI AVGVILVGCCPGGTASNVM TFLAKGNTALSVAVTTISTLLAPVVTPLLIMLFA	151
Qy	150	WSWSLQQNL TIPYQNIGITLV-CLTIPVAFGVYVNYRWPKQ-SKII--LKIGAVVGGVLL	205
		: : :: : : : : : : : :	
Db	152	KEW-----LPVSPGSLFISILQAVLFPIIAGLIVKMFFRKQVAKAVHALPLVSVIG----	202
Qy	206	L V V A V A G V V L A K G S W N --- S D I T L L T I S F I F P L I G H V T G F L L A L F T H Q S W Q R C R T I S L E T	262
		: : :: : :: : :: : :: : ::	
Db	203	-I V A I V S A V V S G N R E N L L Q S G L L I F S V I L H N G I G Y L L G F L C A K L L K M D Y P S Q K A I A I E V	261

```

Qy      263 GAQNIQMCITMLQLSF TAEHLVQMLS FPLA-YGLFQLIDGFLIVAAYQTYKRRLKNKH-G 320
      | |      : | :      | : : | | : : : | : : | | : : : | | |
Db      262 GMQN-----SGLGAALATAHFSPLSAVPSAIFSVWHNLSGSML-ATY--WSKKVKKKQAG 313

Qy      321 KKNS 324
      | : |
Db      314 SKSS 317

```

RESULT 7

D90031

hypothetical protein SA2112 [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C;Accession: D90031

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto, S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: D90031

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-305 <KUR>

A;Cross-references: GB:BA000018; PID:g13702121; PIDN:BAB43413.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: SA2112

C;Superfamily: Bacillus subtilis sodium-dependent transporter yocS

```

Query Match          15.2%;  Score 301.5;  DB 2;  Length 305;
Best Local Similarity 27.9%;  Pred. No. 2.1e-17;
Matches 68;  Conservative 60;  Mismatches 101;  Indels 15;  Gaps 4;

```

```

Qy      41 MMGLIMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAIIV 100
      :|::| :| ::      : | : :|: || :|| |::| || | | | :| |
Db      41 LLGIVMLGMGLTITPNDFKMVFKAPRAVIIGVCLQFSIMPTLAFIIAKSFHLPDPDIAGV 100

Qy     101 LIMCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQONLTI 160
      ::||| ||| ||: ::      :: ||:|:| ||: | : | ||| : | : :
Db     101 ILVGCCPGGTSSNVMSYLAKEANVALSVSITTVSTLLAPFVTPALTYLFANEWLEVSFLSM 160

Qy     161 PYQNIIGITLVCLTIPVAFGV---YVNYRWPQKSKIILKIGAVVGGVLLLVAVAVAG---VV 214
      : : : | ||:| | : : | : : : | | :|| | : | : | | :
Db     161 LWSVVQVVL----IPIALGIVLQIINRKIAEKASTALPIISVVAISLILAIIVVGGSKHQI 216

Qy     215 LAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQNIQMCITML 274
      | |      : : : : : : :|: | |      | : :|:| | | | : : :
Db     217 LTTG-----LLIFLVVILHNVLGYTIGYWLARLLKLDRQDQKAVSIEVGMQNSGLAVSLA 271

Qy     275 QLSF 278
      | |
Db     272 ALHF 275

```

RESULT 8

AD3295

sodium/bile acid cotransporter homolog, sbf family BMEI0346 [imported] -
Brucella melitensis (strain 16M)

C;Species: Brucella melitensis

C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002

C;Accession: AD3295

R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;
Ivanova, N.; Anderson, I.; Bhattacharyya, A.; Lykidis, A.; Reznik, G.;
Jablonski, L.; Larsen, N.; D'Souza, M.; Bernal, A.; Mazur, M.; Goltsman, E.;
Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesson, J.J.; Haselkorn,
R.; Kyrpides, N.; Overbeek, R.

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen Brucella
melitensis.

A;Reference number: AD3252; PMID:11756688

A;Accession: AD3295

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-318 <KUR>

A;Cross-references: GB:AE008917; PIDN:AAL51527.1; PID:g17982244; GSPDB:GN00190

A;Experimental source: strain 16M

C;Genetics:

A;Gene: BMEI0346

A;Map position: I

C;Superfamily: Bacillus subtilis sodium-dependent transporter yocS

Query Match 15.1%; Score 299.5; DB 2; Length 318;
Best Local Similarity 27.9%; Pred. No. 3.2e-17;
Matches 79; Conservative 66; Mismatches 117; Indels 21; Gaps 7;

```
Qy      33 VFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSL 92
      :|      |::|::|| :| :: :      :||: :|:|:| || :|| | ||      :
Db      35 IFAPWIVVLLGIIMFGMGLTISGKDFAEVAKRPFDAIGVLAQFIIMPLLAVLLTRIIPM 94

Qy      93 KPVQAI AVLIMGCCPGGTISNIFTFWVDGDMDSLISMTTCSTVAALGMMPLCIYLYTWSW 152
      | | |::| ||||| ||: | : ||: ||:: | : |: | : : :
Db      95 SPEVAAGVILVGCCPGGTASNVMYTYPKGDVALSVACTSVTLLAPLVTPFLVWFFA--- 151

Qy     153 SLQQNL TIPYQNIGITLV-CLTIPVAFGVYVNYRWP---KQSKIILKIGAVVGGVLLL-- 206
      | | : : :|::| : :|:| | : | | : :| : :||| |:::
Db     152 --SQYLPVDAMSMFISIVKVILVPLALGFVLQKLVPGVVKAAVPMPLVSVVGVIVLIVAA 209

Qy     207 VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQN 266
      |||| :|: | : : : : | :|: | | | : : ||:| | ||
Db     210 VVAVNKAAIAQ----SGLLIFAVVVLHNCCGLLLGYFAARFAGLSLAKRKAISIEVGMQN 265

Qy     267 IQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAY 308
      : | : | : : | | : : : | | | : : |
Db     266 -----SGLGAALANAHFSPLAAVPSAVFSVWHNISGALVASYY 303
```

RESULT 9

B83757

sodium-dependent transporter BH0858 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: B83757

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiram, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis.

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: B83757

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-323 <STO>

A;Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04577.1;

GSPDB:GN00137

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH0858

C;Superfamily: Bacillus subtilis sodium-dependent transporter yocS

Query Match 15.1%; Score 299.5; DB 2; Length 323;
Best Local Similarity 24.8%; Pred. No. 3.2e-17;
Matches 80; Conservative 80; Mismatches 122; Indels 41; Gaps 9;

```
Qy      34 FTVVS---TVMGGLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISE 90
      || :: |::|::|| :| :::: ::::| ||::| ||::| |::| ::|
Db      33 FTWITPHITILLGVIMFGMGLTLKLSDFRIVLQKPIPVLVGVLAQFVIMPLVAFALAYAF 92

Qy      91 SLKPVQAIIVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTW 150
      :| | | | ::::| ||||| ||: : |:: |::||: ||: | : | : |
Db      93 NLPPELAAGLVLVGACPGGTASNVMVYLAKGNVAASVAMTSVSTMLAPIVTPFILLLLAG 152

Qy     151 SWSLQQNLTIPIYQNIQITLV-CLTIPVAFGVYVNYRWPK---QSKIILKIGAVVGGVLLL 206
      | | | : : :::: : :|::| |::| | :| :| : ::| :: :
Db     153 QW-----LPIDAKAMFVSILQMIIVPIALGLFVRKMAPNAVDKSTAVLPLVSIV-AIMAI 206

Qy     207 VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLETGAQN 266
      | || | | : : | : : : | : |::| | | | ||::| | ||
Db     207 VSAVVGANQANLMSGALLFLAV-MLHNVEGLLLGYLTAKFVGLDESTRAISIEVGMQN 265

Qy     267 IQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAYQTYKRRLKNKHGKKNKG 325
      : | : | : : | | : : | | ::|: :
Db     266 -----SGLGAALAGNHFSPLAALPSAIFSVVHNISGPVLVSIWS----- 304

Qy     326 CTEVCHTRKSTSSRETNAFLEVN 348
      : || |::| ::|:
Db     305 -----RSAKSAQKRQSDADMKVD 322
```

RESULT 10

T02645

hypothetical protein At2g26900 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein F12C20.6

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001

C;Accession: T02645; C84666
R;Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.
A;Reference number: Z14685
A;Accession: T02645
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338 <ROU>
A;Cross-references: EMBL:AC005168; NID:g3426033; PID:g3426051
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhagen, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84666
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-338 <STO>
A;Cross-references: GB:AE002093; NID:g3426051; PIDN:AAC32250.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g26900; F12C20.6
A;Map position: 2
A;Introns: 22/2; 61/3; 99/3; 120/3; 163/2; 190/3; 208/1; 240/3; 293/3

Query Match 14.4%; Score 284; DB 2; Length 338;
Best Local Similarity 25.8%; Pred. No. 6.4e-16;
Matches 80; Conservative 54; Mismatches 120; Indels 56; Gaps 8;

```

Qy      9  SACPANSSEEELPVGLEVHGNI-----ELVFTV-----VSTVMM 42
      | |  |:|      | |:|      ||: |:      : |: :
Db      15  SQCRINTSRVVCKAAAGVSGDLPESTPKELSQYEKIIELLTTFLWVWTWLETDLFTLGL 74

Qy      43  GLLMFSLGCSEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLI 102
      | || |:| ::      :| || : || |: : |      :|:|:|: |      | :::
Db      75  GFLMLSMGLTLTFEDFRRCRNPNWTVGVGFLAQYMIKPILGFLIAMTLKLSAPLATGLIL 134

Qy      103  MGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PY 162
      : ||||| ||: |:      |: : ||: |||||: |: | ||      | :|
Db      135  VSCCPGGQASNVATYISKGNVALSVMTCSTIGAIIMTPLL-----KLLAGQLVPV 187

Qy      163  QNIGI---TLVCLTIPVAFGVVNYRWPKQSKIILKIGAVVGGVLLLVVA-----VAGV 213
      |: | : :| || | :|| : |: : :|: | : :      || |
Db      188  DAAGLALSTFQVVLVPTIIGVLANEFFPKFTSKIITVTPPLIGVILTLLCASPIGQVADV 247

Qy      214  VLAKGSWNSDITLLTISFIFPL-IGHVTGFLALFTHQ---SWQRCRTISLETGAQNIQM 269
      : :|:      | |: : | | : : :      |||:| | |: :
Db      248  LKTQGA-----QLILPVALHAAFAIGYWISKFSFGESTSRTISIECGMQSSAL 297

```

```

Qy      270 CITMLQLSFT 279
          : |  ||
Db      298 GFLLAQKHFT 307

```

RESULT 11

F83236

probable transporter PA3264 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: F83236

R; Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: F83236

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-311 <STO>

A:Cross-references: GB:AE004749; GB:AE004091; NID:g9949388; PIDN:AAG06652.1;

GSPDB:GN00131; PASP:PA3264

A; Experimental source: strain PA01

C; Genetics:

A;Gene: PA3264

C;Superfamily: Bacillus subtilis sodium-dependent transporter yocS

Query Match 13.6%; Score 269.5; DB 2; Length 311;
Best Local Similarity 25.7%; Pred. No. 9.2e-15;
Matches 75; Conservative 74; Mismatches 116; Indels 27; Gaps 8;

Qy	32	LVFTTVVSTVMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFS	91
Db	34	LPLTAAIAPLLGLVMFGMLTLKGEDFREVARHPIRVLIGVLAQFVIMPGLAWLLCRLLQ	93
Qy	92	LKPVQAIHAVLIMGCCPGGTISNIFTFWVDGDMDLISISMTCTCSTVAALGMMPLCIYLYTWS	151
Db	94	LPAEIAVGVLVGCCPGGTASNVTWLSRGDVALSVAITSVTTLAPLVTPALVWLLASA	153
Qy	152	WSLQQNLTIPIYQONIGITLV-CLTIPVAFGVYVNYRWPQSKIILKIGAVVGGVLLLVA	210
Db	154	W-----LPVSFAAMFLSILQVVLVPIALGLLAQRLLGERTQVAEVLPLV-SVFSIVVII	207
Qy	211	AGVVLAKGSWNSDITLLTISFIF-----PLIGHVTGFLALFTHQSWQRCRTISLETGA	264
Db	208	AAVVAASQARIAESGLLIMAVVMLHNGFGLLLGYLTKLTKMPLAQR----KALAEVGM	263
Qy	265	QNIQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAYQTYKRRL	315
Db	264	QN-----SGLGAALANAHSPLAAVPSALFSVWHNLSGSLAALF----RRL	306

RESULT 12

B81168

transporter NMB0705 [imported] - *Neisseria meningitidis* (strain MC58 serogroup B)

C;Species: *Neisseria meningitidis*

C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C;Accession: B81168

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Ketchum, K.A.; Hood, D.W.; Peden, J.F.; Dodson, R.J.; Nelson, W.C.; Gwinn, M.L.; DeBoy, R.; Peterson, J.D.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Mason, T.; Ciecko, A.; Parksey, D.S.; Blair, E.; Cittone, H.; Clark, E.B.; Cotton, M.D.; Utterback, T.R.; Khouri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.;

Rappuoli, R.; Venter, J.C.

A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755; PMID:10710307

A;Accession: B81168

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-315 <TET>

A;Cross-references: GB:AE002425; GB:AE002098; NID:g7225930; PIDN:AAF41122.1;

PID:g7225934; GSPDB:GN00119; TIGR:NMB0705

A;Experimental source: serogroup B, strain MC58

C;Genetics:

A;Gene: NMB0705

C;Superfamily: *Bacillus subtilis* sodium-dependent transporter yocS

Query Match 13.5%; Score 266.5; DB 2; Length 315;
Best Local Similarity 26.6%; Pred. No. 1.6e-14;
Matches 81; Conservative 65; Mismatches 102; Indels 57; Gaps 10;

```

Qy      41 MMGLIMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAIIV 100
      ::|::|| :| ::      : | : :|:: || :|| ||::||: :| :| :|
Db      43 LLGIIMFGMGLTLKPSDFDILFKHPKVVIIGVIAQFAIMPATAWLLSKLLNLPAEIAVGV 102

Qy     101 LIMGCCPGGTISNIFTFWDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI 160
      ::|:||||| ||: |: |:: ||::|: ||: : :| |:| :| |
Db     103 ILVGCCPGGTASNVMTYLARGNVALSVAVTSVSTLISPLLTP-AIFLML----AGEMLEI 157

Qy     161 PYQNIGITLV-CLTIPVAFGVVYVNYRWPQSK-----IILKIGAVVG----- 201
      : :::| : :|: |: |: ::: :| :| |||||
Db     158 QAAGMLMSIVKMVLLPIVLGLIVHKVLGSKTEKLTDALPLVSVAIVLIIGAVVGASKGK 217

Qy     202 ----GVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLLALEFTHQSWQRCRT 257
      |:|: | ||| | ||:: || |:| : :|
Db     218 IMESGLLIFAV----VVLHNG-----IGYLLGFFAAKWTGLPYDAQKT 256

Qy     258 ISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLA-YGLFQLIDGFLIVAAYQTYKRRLK 316
      :::| | || : : | | :| :| | : :| | | : ||
Db     257 LTIEVGMQNSGLAAALAAAHFAAAPVV---AVPGALFSVWHNISGSLLA----TYWAAKA 309

Qy     317 NKHGK 321

```

|||
Db 310 GKHKK 314

RESULT 13

E81937

probable transmembrane transport protein NMA0909 [imported] - *Neisseria meningitidis* (strain Z2491 serogroup A)

C;Species: *Neisseria meningitidis*

C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C;Accession: E81937

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morelli, G.; Basham, D.; Brown, D.; Chillingworth, T.; Davies, R.M.; Davis, P.; Devlin, K.; Feltwell, T.; Hamlin, N.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, M.A.; Rutherford, K.M.; Simmonds, M.; Skelton, J.; Whitehead, S.; Spratt, B.G.; Barrell, B.G. Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: E81937

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-315 <PAR>

A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84186.1;

PID:g7379621; GSPDB:GN00124; NMA0909

A;Experimental source: serogroup A, strain Z2491

C;Genetics:

A;Gene: NMA0909

C;Superfamily: *Bacillus subtilis* sodium-dependent transporter yocS

Query Match 13.4%; Score 265.5; DB 2; Length 315;
Best Local Similarity 25.9%; Pred. No. 2e-14;
Matches 79; Conservative 68; Mismatches 101; Indels 57; Gaps 10;

```
Qy      41 MMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI 100
      ::|::| | :| ::      : | :|:: || :||| ::| :| | :|
Db      43 LLGIIMEFGMGLTLKPSDFDILFKHPKAVIIGVIAQFAIMPATAWLLSKLLNLP 102
      ::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      101 LIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQ 160
      ::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      103 ILVGCCPGGTASNVMTYLARGNVALSVAVTSVSTLISPLLTP-AIFLML---- 157
      ::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      161 PYQNIGITLV-CLTIPVAFGVYVNYRWPKQSK-----IILKIGAVVG----- 201
      :: ::| : ::| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      158 QAGSMLMSIVKMVLLPIVLGLIVHKVLGSKTEKLTDALPLVSVAIVLIIGAV 217
      :: ::| : ::| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

Qy      202 ----GVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRC 257
      |::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      218 IMESGLLIFAV----VVLHNG-----IGYLLGFFAAKWTGLPYDAQKT 256
      ::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      258 ISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAYQTYK 316
      ::| | | | : : | : ::| | : ::| | | : | | : | |
Db      257 LAIEVGMQNSGLAAALAAHFA---VAPVAVPGALFSVWHNISGSLLA---TYWA 309
      ::| | | | : : | : ::| | : ::| | | : | | : | |

Qy      317 NKHGK 321
      |||
```

Db 310 GKHKK 314

RESULT 14

D83438

probable transporter PA1650 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: D83438

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: D83438

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-297 <STO>

A;Cross-references: GB:AE004593; GB:AE004091; NID:g9947619; PIDN:AAG05039.1;

GSPDB:GN00131; PASP:PA1650

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA1650

C;Superfamily: *Bacillus subtilis* sodium-dependent transporter yocS

Query Match 13.0%; Score 257.5; DB 2; Length 297;
Best Local Similarity 26.2%; Pred. No. 8.5e-14;
Matches 75; Conservative 60; Mismatches 128; Indels 23; Gaps 7;

```
Qy      33 VFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSL 92
      : | :   : :| :| | | :   : | | : ||| :| | | :| :   : | :|
Db      6 ILTLFLPIALGIIMLGGLSLTPADFLRVVRYPKPVLVGLVCQIVLLPLACFLIVQGFAL 65

Qy      93 KPVQAIAVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSW 152
      :   | : : : :   ||| | :| : :   || : | :| :| : :| | : | | : |
Db      66 EAALAVGMMLLAASPGGTTANLYSHLAHGDVALNITLTAVNSVIAILTMPLIVNL----- 120

Qy     153 SLQ-----QNLTIPTYQNIGITLVCLTIPVAFGVVYNYRWPQSKIILKIGAVVGGVLLL 206
      |||   | : : : :   | :   ||| | : | | | :   : | : : : |||
Db     121 SLQYFMGDGQAIPLQFGKVQVFVIVLGPVAIGMLVRNRLPAVADRLQKPVKILSALLLL 180

Qy     207 VVAVAGVVLAKGSWNSDITLLTI----SFIFPLIGHVTGFLLEFTHQSWQRCRTISLET 262
      | : :   : ||| | : :|   :   : | | :   | : :   :   : | : |
Db     181 VIIL--LALAK-DWQTFVTYAPVVGLAALAFNLLSLAVGYWVPRLLRLPKAQAVAIGMEI 237

Qy     263 GAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGFLIVAAY 308
      | | : | | || : :   | | || :   | || :
Db     238 GIHNGLAIA-LALSPSLNNSTMAIPPAIYGVLM----FFTAAAF 278
```

RESULT 15

E70482

Na(+) dependent transporter (Sbf family) - Aquifex aeolicus
 C;Species: Aquifex aeolicus
 C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jun-2000
 C;Accession: E70482
 R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
 Nature 392, 353-358, 1998
 A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A;Reference number: A70300; MUID:98196666; PMID:9537320
 A;Accession: E70482
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-297 <AQF>
 A;Cross-references: GB:AE000774; NID:g2984324; PIDN:AAC07854.1; PID:g2984333; GB:AE000657
 A;Experimental source: strain VF5
 C;Genetics:
 A;Gene: sbf
 C;Superfamily: Bacillus subtilis sodium-dependent transporter yocS

Query Match 12.6%; Score 250; DB 2; Length 297;
 Best Local Similarity 26.0%; Pred. No. 3.5e-13;
 Matches 81; Conservative 64; Mismatches 126; Indels 40; Gaps 10;

```

Qy      29 NLELVFTVVSTVMMGLL-----MFSLGCSVEIRKLWSHIRRPWGIA 69
      :  :  :||  ::||  |  ||  ::  ||:  :
Db      3 DFSFLILVLSLLGILFPEFFANLKPLILPLLIVIMLSMGLTLTPEDFKEIARKPFIVF 62

Qy      70 VGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISM 129
      |  ||  :  ||  :  |||  |  |  :  |::|  |||  ||:  |  ||:  |||
Db      63 YGALLQYTMPLSGYLLSKLFKLPELLVGVVLVGSAPGGTASNLITYLSRGDLSYSISM 122

Qy     130 TTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPIYQNI-GITLVCLTIPVAFGV----YVNY 184
      ||  ||:  :  ||  |:  :  :  :  :  :  ||  :  :||  |:  :  |
Db     123 TTTSTLLSPLFTPLWITYVLAKY-----VEVPFLSMFETTLKIVIVPVLGMLVRLRYFLRY 177

Qy     185 RWPQSKIIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL 244
      :  |  |  |  ||  :  |::||  :  :|  :  :  |  :  :  |  :  |
Db     178 QINKVEK-FLPFLAVFS--ISLIIAVIFALNSKLLKELSFLVLSVVLHNVLGFLLGYLEF 234

Qy     245 ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFLI 304
      |  :  :  :  :||  |  ||  :  |  :  |  :  :  :  :  |  :  |  ||
Db     235 GLLAGLDKRKVKALSIEVGMQNSGLS-TVLALKYFS----KVSALPSA--IFSLSQN-LI 286

Qy     305 VAAYQTYKRRL 315
      :  |||
Db     287 GVVLSLFFRRL 297

```

Search completed: March 23, 2004, 14:37:49
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:35:53 ; Search time 45 Seconds
(without alignments)
2169.470 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEL.....PGPMDCHRALEPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query				
No.	Score	Match Length	DB	ID		Description

1	1979	100.0	377	13	US-10-091-628-2	Sequence 2, Appli
2	1415	71.5	373	11	US-09-981-151A-40	Sequence 40, Appl
3	1200.5	60.7	326	11	US-09-981-151A-12	Sequence 12, Appl
4	886	44.8	347	11	US-09-981-151A-41	Sequence 41, Appl
5	884	44.7	348	11	US-09-981-151A-45	Sequence 45, Appl
6	871	44.0	348	11	US-09-981-151A-42	Sequence 42, Appl
7	871	44.0	348	11	US-09-981-151A-43	Sequence 43, Appl
8	871	44.0	348	11	US-09-981-151A-44	Sequence 44, Appl
9	553	27.9	349	14	US-10-288-222A-16	Sequence 16, Appl
10	534.5	27.0	440	15	US-10-085-198-114	Sequence 114, App
11	322.5	16.3	367	15	US-10-093-463-22	Sequence 22, Appl
12	322.5	16.3	367	15	US-10-093-463-26	Sequence 26, Appl
13	322.5	16.3	438	13	US-10-091-628-5	Sequence 5, Appli
14	322.5	16.3	438	15	US-10-093-463-24	Sequence 24, Appl
15	322.5	16.3	438	15	US-10-108-260A-3362	Sequence 3362, Ap
16	268.5	13.6	335	9	US-09-738-626-4892	Sequence 4892, Ap
17	265.5	13.4	196	11	US-09-833-245-1622	Sequence 1622, Ap
18	264.5	13.4	324	9	US-09-738-626-6054	Sequence 6054, Ap
19	251	12.7	402	12	US-10-425-114-50020	Sequence 50020, A
20	227	11.5	186	11	US-09-981-151A-82	Sequence 82, Appl
21	171	8.6	417	12	US-10-424-599-253710	Sequence 253710,
22	142	7.2	65	9	US-09-864-761-46433	Sequence 46433, A
23	138	7.0	132	12	US-10-424-599-184632	Sequence 184632,
24	130.5	6.6	350	14	US-10-156-761-8709	Sequence 8709, Ap
25	126	6.4	716	12	US-10-282-122A-69161	Sequence 69161, A
26	120	6.1	716	12	US-10-282-122A-76294	Sequence 76294, A
27	119	6.0	701	12	US-10-282-122A-56094	Sequence 56094, A
28	117.5	5.9	461	12	US-10-282-122A-51204	Sequence 51204, A
29	117	5.9	721	12	US-10-282-122A-42827	Sequence 42827, A
30	115	5.8	582	12	US-10-282-122A-59988	Sequence 59988, A
31	114	5.8	372	9	US-09-966-871-80	Sequence 80, Appl
32	114	5.8	372	12	US-10-458-860-80	Sequence 80, Appl
33	114	5.8	372	13	US-10-039-645-80	Sequence 80, Appl
34	114	5.8	372	15	US-10-328-916-10	Sequence 10, Appl
35	114	5.8	688	12	US-10-282-122A-68097	Sequence 68097, A
36	113.5	5.7	772	9	US-09-935-799A-2	Sequence 2, Appli
37	113.5	5.7	772	9	US-09-935-799A-5	Sequence 5, Appli
38	113	5.7	628	12	US-10-282-122A-69938	Sequence 69938, A
39	108.5	5.5	140	12	US-10-424-599-183649	Sequence 183649,
40	107.5	5.4	693	14	US-10-225-567A-518	Sequence 518, App
41	107.5	5.4	693	14	US-10-073-054-2	Sequence 2, Appli
42	107.5	5.4	693	15	US-10-295-027-768	Sequence 768, App
43	107.5	5.4	693	15	US-10-436-715-41	Sequence 41, Appl
44	106	5.4	695	16	US-10-333-946-13	Sequence 13, Appl
45	105	5.3	324	9	US-09-738-626-6866	Sequence 6866, Ap

ALIGNMENTS

RESULT 1
 US-10-091-628-2
 ; Sequence 2, Application US/10091628
 ; Publication No. US20020164627A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Nepomnichy, Boris
; APPLICANT: Burnett, Michael B.
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020164627A1el Human Transporter Proteins and
Polynucleotides Encoding the
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0314-USA
; CURRENT APPLICATION NUMBER: US/10/091,628
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,009
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/284,152
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-091-628-2

Query Match 100.0%; Score 1979; DB 13; Length 377;
Best Local Similarity 100.0%; Pred. No. 4.2e-181;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWS	60
Db	1	MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWS	60
Qy	61	HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPQAIHAVLIMGCCPGGTISNIFTFWVD	120
Db	61	HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPQAIHAVLIMGCCPGGTISNIFTFWVD	120
Qy	121	GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTYQNIGITLVCLTIPVAFGV	180
Db	121	GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTYQNIGITLVCLTIPVAFGV	180
Qy	181	YVNYRWPKQSKIILKIGAVVGGVLLLVAAGVVLAKGSWNSDITLLTISFIFPLIGHVT	240
Db	181	YVNYRWPKQSKIILKIGAVVGGVLLLVAAGVVLAKGSWNSDITLLTISFIFPLIGHVT	240
Qy	241	GFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLID	300
Db	241	GFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLID	300
Qy	301	GFLIVAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP	360
Db	301	GFLIVAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP	360
Qy	361	MDCHRALEPVGHITSCE	377
Db	361	MDCHRALEPVGHITSCE	377

RESULT 2
US-09-981-151A-40
; Sequence 40, Application US/09981151A

; Publication No. US20030212256A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gerlach, Valerie
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Muriel M
; APPLICANT: Smithson, Glennnda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A
; APPLICANT: Stone, David J
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Shimkets, Richard A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Guo, Xiaojia
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Burgess, Catherine E
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Gorman, Linda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-168
; CURRENT APPLICATION NUMBER: US/09/981,151A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,040
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,058
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-981-151A-40

Query Match 71.5%; Score 1415; DB 11; Length 373;
Best Local Similarity 70.3%; Pred. No. 5e-127;
Matches 265; Conservative 50; Mismatches 58; Indels 4; Gaps 2;

```
Qy      1 MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVSTVMMGLLMFSLGCSVEIRKLWS 60
      | :|: :| || ||:|:| ||:| | ||:|:|:| ||:|:| || |||| :|||
Db      1 MSTDCAGNSTCPVNSTEEDPPVGMEGHANLKLFTVLSAVMVGLVMFSFGCSVESQKLWL 60

Qy     61 HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD 120
      |:||||| ||||| ||||| | || | |||||:| |||||: |||||
Db     61 HLRRPWGIAVGLLSQFGLMPLTAYLLAIGFGLKPFQAI AVLMMGSCPGGTISNVLTFWVD 120

Qy    121 GDMDLISISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPIYQNIGITLVCLTIPVAFGV 180
      ||||| ||||| ||||| |||||:|:| ||:| || | ||||:|||| | :||| ||
Db    121 GDMDLISISMTTCSTVAALGMMPLCLYIYTRSWTLTQNLVIPYQSIGITLVSLVVPVASGV 180

Qy    181 YVNYRWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT 240
      ||||| ||||: :||:|:|:|:| ||||| |:|||| ||:|:| || | ||||:||||
Db    181 YVNYRWPKQATVILKVGAILGGMLLLVAVTGMVLAKG-WNTDVTLLVISCIFFPLVGHVT 239

Qy    241 GFLALFLTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLID 300
      |||| | ||||| |||||:|||||:| | ||||:|:|:|:|:| | |||||:|
Db    240 GFLAFLTHQSWQRCRTISLETGAQNIQLCIAMLQLSFSAEYLVQLLNFALAYGLFQVLH 299

Qy    301 GFLIVAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
      | ||||| |||| |:| :: | :||: :: ||:|:|: |:| |:| || |
Db    300 GLLIVAAYQAYKRRQKSKCRRQHPDCPDVCYEKQ---PRETSAFLDKGDEAAVTLGPVQP 356

Qy    361 MDCHRALEPVGHITSCE 377
      ||| | || |||
Db    357 EQHHRAAELTSHIPSCE 373
```

RESULT 3

US-09-981-151A-12

; Sequence 12, Application US/09981151A

; Publication No. US20030212256A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R

; APPLICANT: Gerlach, Valerie

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Muriel M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Millet, Isabelle

; APPLICANT: Peyman, John A

; APPLICANT: Stone, David J

; APPLICANT: Gunther, Erik

; APPLICANT: Ellerman, Karen

; APPLICANT: Shimkets, Richard A

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Guo, Xiaojia

; APPLICANT: Patturajan, Meera

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Burgess, Catherine E

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly A

```

; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Gorman, Linda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-168
; CURRENT APPLICATION NUMBER: US/09/981,151A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,040
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,058
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-981-151A-12

```

```

Query Match          60.7%; Score 1200.5; DB 11; Length 326;
Best Local Similarity 80.6%; Pred. No. 1.5e-106;
Matches 250; Conservative 11; Mismatches 30; Indels 19; Gaps 6;

```

```

Qy      1 MRANCSSSSACPANSSEEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWS 60
        |||
Db      1 MRANCSSSSACPANSSEEELPVGLEVHGNLELVFTVVSTIMMGLLMFSLGCSVEIRKLWS 60

Qy      61 HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD 120
        |||
Db      61 HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCRG---APSLTFSPS 117

Qy      121 GDMDLSISM TTCSTVAALGMMPLCIYLYTWSWSLQQNL TIPYQNI----GITLVCLTIPV 176
        | |:: ||
Db      118 GLMEIWIS-----GALGMMPLCIYLYTWSWSLQQNL TIPYQNIGLSLGITLVCLTIPV 170

Qy      177 AFGVYVNYRWPQSKII LKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLI 236
        |||
Db      171 AFGVYVNYRWPQSKII LK--AVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLI 228

Qy      237 GHVTGFL LALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSF-PLAYGL 295

```

		: :	: ::	: :
Db	229	GHVTGFLALFTHQSWQ--RTLPIFLGLAFKTPCDTLLAMTSCPECSRLIYAFIPLLYGL	286	
QY	296	FQLIDGFLIV 305		
Db	287	FQLIDGFLIV 296		

RESULT 4

US-09-981-151A-41

; Sequence 41, Application US/09981151A

; Publication No. US20030212256A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R

; APPLICANT: Gerlach, Valerie

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Muriel M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Millet, Isabelle

; APPLICANT: Peyman, John A

; APPLICANT: Stone, David J

; APPLICANT: Gunther, Erik

; APPLICANT: Ellerman, Karen

; APPLICANT: Shimkets, Richard A

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Guo, Xiaojia

; APPLICANT: Patturajan, Meera

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Burgess, Catherine E

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Gangolli, Esha A

; APPLICANT: Fernandes, Elma R

; APPLICANT: Gorman, Linda

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-168

; CURRENT APPLICATION NUMBER: US/09/981,151A

; CURRENT FILING DATE: 2001-10-16

; PRIOR APPLICATION NUMBER: 60/241,040

; PRIOR FILING DATE: 2000-10-17

; PRIOR APPLICATION NUMBER: 60/241,058

; PRIOR FILING DATE: 2000-10-17

; PRIOR APPLICATION NUMBER: 60/241,063

; PRIOR FILING DATE: 2000-10-17

; PRIOR APPLICATION NUMBER: 60/241,243

; PRIOR FILING DATE: 2000-10-17

; PRIOR APPLICATION NUMBER: 60/242,152

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/242,482

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/242,611

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/242,612

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/242,880

; PRIOR FILING DATE: 2000-10-24

```
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Oryzctolagus cuniculus
US-09-981-151A-41
```

```
Query Match          44.8%; Score 886; DB 11; Length 347;
Best Local Similarity 47.0%; Pred. No. 2.4e-76;
Matches 164; Conservative 73; Mismatches 98; Indels 14; Gaps 5;
```

```
Qy      11 CPANSS--EEELPVGLEVHGN--LELVFTTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPW 66
      | ||:: | | : | | : | |:: |::|::|::|::| | ||||
Db      8 CLANATVCEGASCVAPESNFNAILS SVLSTVLTILLALVMFMSMGCNVEIKKFLGHIRRPW 67

Qy      67 GIAVGLLCQFGLMPFTAYLLAISFSLKPQVQAI AVLIMGCCPGGTISNIFTFWVDGDMDL 126
      || :| |||||::|| | ::||::| : |::| | ||||| || | :||| |||||
Db      68 GIFIGFLCQFGIMPLTGFVLAVAFGIMPIQAVVVLIMGCCPGGTASNILAYWVDGDMDL 127

Qy      127 ISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVVNYRW 186
      :|||::|: |||||::|::| | : || || | :|| | :||: |::|::|
Db      128 VSMTTCSTLLALGMMPLCLYVYTKMWVDSGTIVIPYDNIGTSLVALVVPVSIGMEFVNHW 187

Qy      187 PKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLAL 246
      |::| |||||::|::| | ||::| | :| : | : | | |||: | : || ||
Db      188 PQKAKIILKVGSIAGAVLIVLIAVVGILYQSAWII EPKLWII GTIFPMAGYSLGFFLAR 247

Qy      247 FTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPLAYGLFQLIDGFLIVA 306
      | | ||||:|||| | | :| |::|||: | | : :|| | :||: : :
Db      248 IAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLTYVTFPLIYSIFQIAFAAIFLG 307

Qy      307 AYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETN AFLEVNEEGAITP 355
      | |:: ||| :: : : | | : :| ::| | |
Db      308 IYVAYRK----CHGKNDAEFPDI----KDTKTEPESSFHQMN--GGFQP 346
```

RESULT 5

```
US-09-981-151A-45
; Sequence 45, Application US/09981151A
; Publication No. US20030212256A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gerlach, Valerie
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Muriel M
; APPLICANT: Smithson, Glennnda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A
; APPLICANT: Stone, David J
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Shimkets, Richard A
; APPLICANT: Padigaru, Muralidhara
```

```

; APPLICANT: Guo, Xiaojia
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Burgess, Catherine E
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Gorman, Linda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-168
; CURRENT APPLICATION NUMBER: US/09/981,151A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,040
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,058
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-981-151A-45

```

```

Query Match          44.7%; Score 884; DB 11; Length 348;
Best Local Similarity 46.9%; Pred. No. 3.7e-76;
Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

```

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | : : | : | | :| : | |::: |:|::|::|::| | | :
Db      3 NSSICNPNATICEGDSCIAPESNFNAILSVMSTVLTILLALVMFSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| | | ||||:| | | :|:::| : | |||: | | | ||||| | | | :||| |
Db     63 RRPWGI VVGFLCQFGIMPLTGFVLSVAFGILPVQAVVLIQGCGPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
      |||:| |||||: | |||||:::| | | : ||| :| | :| | | | | :| |

```

[illegible]

RESULT 6

US-09-981-151A-42

; Sequence 42, Application US/09981151A

; Publication No. US20030212256A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R

; APPLICANT: Gerlach, Valerie

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Muriel M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Millet, Isabelle

; APPLICANT: Peyman, John A

; APPLICANT: Stone, David J

```
; APPLICANT:  Gunther, Erik
```

; APPLICANT: Ellerman, Karen

; APPLICANT: Shimkets, Richard A

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Guo, Xiaojia

; APPLICANT: Patturajan, Meera

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Burgess, Catherine E

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Gangolli, Esha A

; APPLICANT: Fernandes, Elma R

; APPLICANT: Gorman, Linda

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-168

; CURRENT APPLICATION NUMBER: US/09/981,151A

; CURRENT FILING DATE: 2001-10-16

; PRIOR APPLICATION NUMBER: 60/241,040

; PRIOR FILING DATE: 2000-10-17

; PRIOR APPLICATION NUMBER: 60/241,058

; PRIOR FILING DATE: 2000-10-17

: PRIOR APPLICATION NUMBER: 60/241,063

; PRIOR FILING DATE: 2000-10-17

: PRIOR APPLICATION NUMBER: 60/241,243

: PRIOR FILING DATE: 2000-10-17

: PRIOR APPLICATION NUMBER: 60/242,152

: PRIOR FILING DATE: 2000-10-20

: PRIOR APPLICATION NUMBER: 60/242.482


```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-981-151A-42
```

```
Query Match          44.0%; Score 871; DB 11; Length 348;
Best Local Similarity 47.1%; Pred. No. 6.6e-75;
Matches 165; Conservative 70; Mismatches 105; Indels 10; Gaps 3;
```

```
Qy      7 SSSACPANSSEEEELPVGLEVHGN---LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | | | | | : | |::: ::|||:|:| | | ||
Db      3 NSSVCSNPATFCEGDSCLVTESNFNAILSTVMSTVLTILLAMVMFMSMGCNVEINKFLGHI 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      :|||| | | ||||:| | :|::: : ||||: ||||| || | ||: |||
Db     63 KRPWGIFVGFLCQFGIMPLTGFILSVASGILPVQAVVVLIMGCCPGGTGSNILAYWIDGD 122

Qy    123 MDLSISMTCSTVAALGMMPLCIYLYTWSWSLQONLTIPYQNIGITLVCLTIPVAFGVYV 182
      ||||:|||||: ||||| ||::| | | : ||| :||:| | | ||: |:|
Db    123 MDLSVSMTCSTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGISLVALVIPVSIGMFV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:| |:|:| ||||| |:| :|:::| | :| : :| : | | ||: |: ||
Db    183 NHKWPQAKIILKIGSIAGAILIVLIAVVG GILYQSAWIIIEPKLWIIIGTIFPIAGYSLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGF 302
      || | | ||||:|||| || |:| |:| |||: | | : :|| | :||:
Db    243 FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA 302

Qy    303 LIVAAYQTYKRRLKNKHGKKN SGCTEVCHTRKS--TSSRETNAFLEVNEE 350
      |:| | |||: ||| :: | | :||| : :|:
Db    303 IILGMYVTYKK----CHGKNDAEFLEKTDNDMDPMPSPFQETNKGFPDEK 348
```

RESULT 7

US-09-981-151A-43

```
; Sequence 43, Application US/09981151A
; Publication No. US20030212256A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gerlach, Valerie
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Muriel M
; APPLICANT: Smithson, Glenda
```

```

; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A
; APPLICANT: Stone, David J
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Shimkets, Richard A
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Guo, Xiaojia
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Burgess, Catherine E
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Gorman, Linda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-168
; CURRENT APPLICATION NUMBER: US/09/981,151A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,040
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,058
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-981-151A-43

```

```

Query Match          44.0%; Score 871; DB 11; Length 348;
Best Local Similarity 47.4%; Pred. No. 6.6e-75;
Matches 167; Conservative 74; Mismatches 97; Indels 14; Gaps 5;

```

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
        :|| || |: | : | | : | | : | |::: ::|||:|:|:| | ||
Db      3 NSSVCPNATVCEGDCSVVPESNFNAILNTVMSTVLTILLAMVMFSGMGCNVEVHKFLGHI 62

```

Qy 63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
 :|||| | | ||||:| | :|::: : ||||: ||||| || | :|:| |
 Db 63 KRPWGIFVGFCLCQFGIMPLTGFI LSVASGILPVQAVVVLIMGCCPGGTGSNILAYWIDGD 122
 Qy 123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYV 182
 |||:|||||: |||||::| | : ||| :||:| | | ||:|:| |
 Db 123 MDLSVSMTCSTLLALGMMPLCLFVYTKMWVDSGTIVIPYDSIGISLVALVIPVSFGMFV 182
 Qy 183 NYRWPKQSKIILKIGAVGGVLLL VVAVAGVV LAKGSWNSDITLLTISFIFPLIGHVTGF 242
 |:|:|:|:|:|:|:| | :|:|:| | | : | : | | | : | |
 Db 183 NHKWPQKAKIILKIGSITGVILIVLIAVIGGILYQSAWIIEPKLWIIGTIFPIAGYSLGF 242
 Qy 243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPLAYGLFQLIDGF 302
 | | | | | :| | | | | :| | | | :| | | :| | | :| | | :| | |
 Db 243 FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA 302
 Qy 303 LIVAAYQTYKRRLKNKHGKNSGCTEVCHTRKSTSSR---ETNAFLEVNEE 350
 :|: | | : :| : | | | :|:
 Db 303 VILGIYVTYRK----CYGKNDAEFLE--KTDNEMDSRPSFDETNGFQPD EK 348

RESULT 8

US-09-981-151A-44

; Sequence 44, Application US/09981151A
 ; Publication No. US20030212256A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: MacDougall, John R
 ; APPLICANT: Malyankar, Muriel M
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Peyman, John A
 ; APPLICANT: Stone, David J
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Shimkets, Richard A
 ; APPLICANT: Padigar, Muralidhara
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Taupier Jr, Raymond J
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Zerhusen, Bryan D
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Spytek, Kimberly A
 ; APPLICANT: Gangolli, Esha A
 ; APPLICANT: Fernandes, Elma R
 ; APPLICANT: Gorman, Linda
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-168
 ; CURRENT APPLICATION NUMBER: US/09/981,151A
 ; CURRENT FILING DATE: 2001-10-16
 ; PRIOR APPLICATION NUMBER: 60/241,040
 ; PRIOR FILING DATE: 2000-10-17
 ; PRIOR APPLICATION NUMBER: 60/241,058
 ; PRIOR FILING DATE: 2000-10-17

```

; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 348
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-981-151A-44

```

```

Query Match          44.0%;   Score 871;   DB 11;   Length 348;
Best Local Similarity 47.4%;   Pred. No. 6.6e-75;
Matches 167;   Conservative 74;   Mismatches 97;   Indels 14;   Gaps 5;

```

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| || |:: | : | | : | | : | |::: ::|||:|:|:| | ||
Db      3 NSSVCPPNATVCEGDSCVVPESNFNAILNTVMSTVLTILLAMVMFSMGCNVEVHKFLGHI 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      :|||| | | ||||:| | :|:|: : |||: ||||| ||| | :|:| |
Db     63 KRPWGIFVGFCLCQFGIMPLTGFI LSVASGILPVQAVVVLIMGCCPGGTGSNILAYWIDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPTYQNIGITLVCLTIPVAFGVYV 182
      |||:| |||||: |||||:|:| | | : || :|:| | | ||:|:| |
Db    123 MDLSVSMTTCSTLLALGMMPLCLFVYTKMWVDSGTIVIPYDSIGISLVALVIPVSFGMFV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |:|:|:|:| |||||:| :|:|:| | :| : :| : | | ||:| : ||
Db    183 NHKWPQKAKIILKIGSITGVILIVLIAVIGGILYQSAWII EPKLWIIGTIFPIAGYSLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGF 302
      || | | |||:| || | | :| |:| | | | : || | :|:| :
Db    243 FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA 302

Qy    303 LIVAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSR---ETNAFLEVNEE 350
      |:| | |:| : || :| | | || || | :|:|
Db    303 VILGIYVTYRK----CYGKNDAEFLE--KTDNEMDSRPSFDETNGKGFQPDEK 348

```

RESULT 9

US-10-288-222A-16

; Sequence 16, Application US/10288222A

```
; Publication No. US20030119742A1
; GENERAL INFORMATION:
; APPLICANT: Logan, Thomas Joseph
; APPLICANT: Galvin, Katherine
; APPLICANT: Chun, Miyoung
; TITLE OF INVENTION: Methods and Compositions to treat
; TITLE OF INVENTION: Cardiovascular Disease Using 139, 258, 1261, 1486, 2398,
2414, 7660, 8587,
; TITLE OF INVENTION: 10183, 10550, 12680, 17921, 32248, 60489 OR 93804
; FILE REFERENCE: MPI2001-286PlR(M)
; CURRENT APPLICATION NUMBER: US/10/288,222A
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-288-222A-16
```

```
Query Match          27.9%; Score 553; DB 14; Length 349;
Best Local Similarity 36.0%; Pred. No. 2.1e-44;
Matches 124; Conservative 77; Mismatches 109; Indels 34; Gaps 10;
```

```
Qy      31 ELVFTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISF 90
      :| :|: |: :| |||:| | :|: :| |:|: |: |:|:| | :| |
Db      24 DLALSVILVFMLFFIMLSLGCTMEFSKIKAHLWKPKGLAIALVAQYGIMPLTAFVLGKVF 83

Qy      91 SLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYT- 149
      || :|:|:|: || ||| :|:|: : |||:| | | | | | | | | | :|:|
Db      84 RLKNIEALAILVCGCSPGGNLSNVFSLAMKGMNLSIVMTTCSTFCALGMMPLLLYIYSR 143

Qy     150 --WSWSLQQNLTIPIYQNIGITLVCLTIPVAFGVYVNYRWPKQSKIILKIGAVVGGVLLLV 207
      : |: :|:| | |:| : || |: : : |: : :| | : : :|:
Db     144 GIYDGDLDKDK--VPYKGVISLVVLIPCTIGIVLKSRRPQYMRVVIKGGMII----ILL 197

Qy     208 VAVAGVVLAKGSWNNDI-----TLTISFIFPLIGHVTGELL-ALFTHQSWQRC-RTIS 259
      :|| ||: : | |: | : | || : |:| ||| | | | |
Db     198 CSVAVTVLSAINVGKSIMFAMTPLLIIATSSLMPIFIGFLLGYVLSALFCLNG--RCRRTVS 255

Qy     260 LETGAQNIQMCITMLQLSFTHAEHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKRRLKNKH 319
      :||| ||:|:| |:| :| | : : ||| | :||| :| |:| : |: : |
Db     256 METGCQNVQLCSTILNVAFPPEVIGPLFFFPLLYMIFQLGEGLLLIAIFWCYE-KFKTPK 314

Qy     320 GKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGPMDC 363
      | | : :| :| | | : | | : | |
Db     315 DK-----TKMIYTAAT'-----EETIPGALGNGTYKGEDC 344
```

RESULT 10

US-10-085-198-114

```
; Sequence 114, Application US/10085198
; Publication No. US20040009907A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-279
```

```

; CURRENT APPLICATION NUMBER: US/10/085,198
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/271,646
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/276,401
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,981
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,858
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/271,840
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/277,324
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/286,096
; PRIOR FILING DATE: 2001-04-21
; PRIOR APPLICATION NUMBER: 60/299,695
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/315,614
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/272,405
; PRIOR FILING DATE: 2001-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 114
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-198-114

```

```

Query Match          27.0%; Score 534.5; DB 15; Length 440;
Best Local Similarity 37.3%; Pred. No. 1.7e-42;
Matches 112; Conservative 60; Mismatches 113; Indels 15; Gaps 5;

```

```

Qy      27 HGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLL 86
      ||   |   :   |:|   |||:|:   :|:| |   :   |||||:|   |:|
Db      103 HGLNVFVGAAALCITMLG-----LGCTVDVNHFGAHVRRPVGALLAALCQFGLLPLLAFL 157

Qy      87 AISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIY 146
      |:| |   | |:|:|: | ||||| :|: :   |||||:| |   ||: || :|:|:|:
Db      158 ALAFKLDEVA AVALLCGCCPGGNLSNLMSLLVDGDMNLSIIMTISSTLLALVLMPLCLW 217

Qy      147 LYTWSW---SLQQNLTIPIYQNI GITLVCLTIPVAFGVYVNYRWPQSKIILKIGAVVGGV 203
      :|:|:|   : |   :|   : :||   ||: ||:| :|: : :   |:|: |
Db      218 IYSWAWINTPIVQ--LLPLGTVTLTLCSTLPIGLGVFI RYKYSRVADYIVKVRPVSLWS 275

Qy      204 LLLVVAV----AGVVLA KGSWNS-DITLLTISFIFPLIGHVTGFL LALFTHQSWQRCRTI 258
      ||: : |   | :|   |   : |:   || |: :|: ||   |   ||:
Db      276 LLVTLVVLFI MTGTMLGP ELLASIPAAVYVIAIFMPLAGYASGYGLATL FHLPPNCKRTV 335

Qy      259 SLETGAQNIQMCITMLQLSFTA EHLVQMLS FPLAYGLFQLIDGFLIVAAYQTYKRRLKNK 318
      ||||:|:|:|   :|:|:|   : : |   ||| | |||   : : |   |:|   : :|
Db      336 CLETGSQNVQLCTAILKLA FPPQFIGSMYMFPLLYALFQSAEAGIFVLIYKMYGSEMLHK 395

```

RESULT 11

US-10-093-463-22

; Sequence 22, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennnda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic
Polypeptides, Nucleic Acids
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424

```
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-22
```

```
Query Match          16.3%; Score 322.5; DB 15; Length 367;
Best Local Similarity 26.9%; Pred. No. 2.8e-22;
Matches 79; Conservative 71; Mismatches 131; Indels 13; Gaps 5;
```

```
Qy      26 VHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYL 85
      :| : :: :: :: | || :|:: : :|| : :| : || ||| :|
Db      66 MHIDRNILMLILPLILLNKCAF--GCKIELQLFQTVWKRLPVILGAVTQFFLMPFCGFL 123

Qy      86 LAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTCSTVAALGMMPLCI 145
      |: :| || |: : ||| :| :||| |:| || ||: || |||:
Db      124 LSQIVALPEAQAFGVVMTCTCPGGGGGYLFALLLDGDFTLAILMTCTSTLLALIMPVNS 183

Qy      146 YLYTWSWSLQQNLTIPTQYQNGITLVCLTIPVAFGVVYNYRWPKQSKIILKIGAVVGGVLL 205
      |:| : | || | ||: :||: |: : :| |: : :| : :|:
Db      184 YIYSRILGLSGTFHIPVSKIVSTLLFILVPVSIGIVIKHRIPEKASFLEIRIRPLSFILM 243

Qy      206 LV----VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLE 261
      | :| | : : : : :| :| :|: | :|: :|
Db      244 FVGIYLTFTVGLVFLK---TDNLEVILLGLLVPALGLLFGYSFAKVCTLPLPVCKTVAIE 300

Qy      262 TGAQNIQMCITMLQLSF--TAEHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKR 313
      :| | : : :||| : :| : | :| : : ||: |: ||
Db      301 SGMLNSFLALAVIQLSFPQSKANLASVAPFTVA--MCSGCEMLLIILVYKAKKR 352
```

RESULT 12

US-10-093-463-26

```
; Sequence 26, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
```


; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic
Polypeptides, Nucleic Acids
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198

```
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-26
```

```
Query Match          16.3%; Score 322.5; DB 15; Length 367;
Best Local Similarity 26.9%; Pred. No. 2.8e-22;
Matches 79; Conservative 71; Mismatches 131; Indels 13; Gaps 5;
```

```
Qy      26 VHG NLELVFTV VSTVMMGLLMFSLGCSVEIRKLW SHIRRPWGIAVGLLCQFGLMPFTAYL 85
      :| :  ::  ::  ::  |  || :|| :  :| : || ||| :|
Db      66 MHIDRNILMLLILPLILLNKCAF--GCKIELQLFQTVWKRPLPVILGAVTQFFLMPFCGFL 123

Qy      86 LAISFSLKPVQAI AVLMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCI 145
      |:  :|  ||  |::  |||  :|  :||  |:|  ||  ||:  ||  ||:
Db      124 LSQIVALPEAQAFGVMTCTCPGGGGGYLFALLLDGDFTLAILMTCTSTLLALIMMPVNS 183

Qy      146 YLYTWSWSLQQNLTIPIYQNI GITLVCLTIPVAFGVVYNYRWPKQSKIILKIGAVVGGVLL 205
      |:|:  |  ||  |  ||:  :  ||:  |:  :  :|  |::  :  :|  :  :|:
Db      184 YIYSRILGLSGTFHIPVSKIVSTLLFILVPV SIGIVIKHRIPEKASFLERIIRPLSFILM 243

Qy      206 LV----VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL LALFTHQSWQRCRTISLE 261
      |  |  |:|  |  ::  ::  :  :  |  :|  :  |:  |  |:|::|
Db      244 FVG IYLTFTVGLVFLK---TDNLEVILLGLLVPALGLLFGYSFAKVCTLPLPVCKTVAIE 300

Qy      262 TGAQNIQMCITMLQLSF--TAEHLVQMLSFP LAYGLFQLIDGFLIVAAYQTYKR 313
      :|  |  :  :  :|||  :  :|  :  |  :  :  :  ||:  |:  ||
Db      301 SGMLNSFLALAVIQLSF PQSKANLASVAPFTVA--MCSGCEMLLIILVYKAKKR 352
```

RESULT 13

US-10-091-628-5

```
; Sequence 5, Application US/10091628
; Publication No. US20020164627A1
; GENERAL INFORMATION:
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Burnett, Michael B.
; APPLICANT: Hu, Yi
```

```
; TITLE OF INVENTION: No. US20020164627A1el Human Transporter Proteins and
Polynucleotides Encoding the
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0314-USA
; CURRENT APPLICATION NUMBER: US/10/091,628
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,009
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/284,152
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-091-628-5
```

```
Query Match          16.3%; Score 322.5; DB 13; Length 438;
Best Local Similarity 26.9%; Pred. No. 3.5e-22;
Matches 79; Conservative 71; Mismatches 131; Indels 13; Gaps 5;
```

```
Qy      26 VHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYL 85
      :| : :: :: :: | || :|:: : :| : :| : || ||| :|
Db      137 MHIDRNILMLILPLILLNKCAF--GCKIELQLFQTVWKRPLPVILGAVTQFFLMPFCGFL 194

Qy      86 LAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCI 145
      |: :| || |:: |||| :| :||| |:| || ||: || |||:
Db      195 LSQIVALPEAQAFGVVMTCTCPGGGGGYLFALLLDGDFTLAILMTCTSTLLALIMMPVNS 254

Qy      146 YLYTWSWSLQQNLTIPIYQNGITLVCLTIPVAFGVYVNYRWPKQSKIILKIGAVVGGVLL 205
      |:|: | || | ||: : :||: |: : :| |::: : :| : :|:
Db      255 YIYSRILGLSGTFHIPVSKIVSTLLFILVPVSI GIVIKHRIPEKASFLEIRIIRPLSFILM 314

Qy      206 LV----VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFL LALFTHQSWQRCRTISLE 261
      | | :| | : : : : :| :| : |: | | :|:::|
Db      315 FVGIIYLTFTVGLVFLK---TDNLEVILLGLLV PALGLLFGYSFAKVCTLPLPVCKTVAIE 371

Qy      262 TGAQNIQMCITMLQLSF--TAEHLVQMLSFPLAYGLFQLIDGFLIVAAYQTYKR 313
      :| | : : :||| : :| : | :| : : ||: |: ||
Db      372 SGMLNSFLALAVIQLSF PQSKANLASVAPFTVA--MCSGCEMLLIILVYKAKKR 423
```

RESULT 14

US-10-093-463-24

```
; Sequence 24, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
```

; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypeptides, Nucleic Acids
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18

```
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 438
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-093-463-24
```

Qy	26	VHGNLELVFTTVVSTVMMGLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYL	85
Db	137	MHIDRNILMLILPLILLNKCAF--GCKIELQLFQTVWKRPLPVILGAVTQFFLMPFCGFL	194
Qy	86	LAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMDSLISMTTCSTVAALGMMPLCI	145
Db	195	LSQIVALPEAQAFGVVMTCTCPGGGGGYLFALLLDGDFTLAILMTCTSTLLALIMMPVNS	254
Qy	146	YLYTWSWSLQQNLTIPIYQNI GITLVCLTIPVAFGVYVYRWPQSKIILKIGAVVGGVLL	205
Db	255	YIYSRILGLSGTFHIPVSKI VSTLLFILVPV SIGIVIKHRIPEKASFLEIRIPLSFILM	314
Qy	206	LV---VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLE	261
Db	315	FVGIYLTFTVGLVFLK---TDNLEVILLGLLVPALGLLFGYSFAKVCTLPLPVCKTVAIE	371
Qy	262	TGAQNIQMCITMLQLSF--TAEHLVQMLSFPPLAYGLFQLIDGFLIVAAAYQTYKR	313
Db	372	SGMLNSFLALAVIQLSFQSKANLASVAPFTVA--MCSGCEMLLIILVYKAKKR	423

RESULT 15

Query Match 16.3%; Score 322.5; DB 15; Length 438;
 Best Local Similarity 26.9%; Pred. No. 3.5e-22;
 Matches 79; Conservative 71; Mismatches 131; Indels 13; Gaps 5;

```

Qy      26 VHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAGVLLCQFGLMPFTAYL 85
      :| : :: :: :: | || :| : : || : :| : || ||| :|
Db      137 MHIDRNILMLILPLILLNKCAF--GCKIELQLFQTVWKRPLPVILGAVTQFFLMPFCGFL 194

Qy      86 LAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCI 145
      |: :| || |:: ||| :| :|| |:| || |:| |||:
Db      195 LSQIVALPEAQAFGVMTCTCPGGGGGYLFALLLDGDFTLAILMTCTSTLLALIMPVNS 254

Qy      146 YLYTWSWSLQQNLTIPTQYQNI GITLVCLTIPVAFGVYVNYRWPKQSKIILKIGAVVGGVLL 205
      |:|: | || | ||: :||: |: : :| |:: : :| : :|
Db      255 YIYSRILGLSGTFHIPVSKI VSTLLFILVPV SIGIVIKHRIPEKASFLEIRIIRPLSFILM 314

Qy      206 LV---VAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLE 261
      | :| | :: : : | :| : |: | :|:|:|
Db      315 FVGIIYLTFTVGLVFLK---TDNLEVILLGLLVPALGLLFGYSFAKVCTLPLPVCKTVAIE 371

Qy      262 TGAQNIQMCITMLQLSF--TAEHLVQMLSFPPLAYGLFQLIDGFLIVAAYQTYKR 313
      :| | : : ::||| : :| : | :| : : ||: |: ||
Db      372 SGMLNSFLALAVIQLSFPQSKANLASVAPFTVA--MCSGCEMLLIILVYKAKKR 423

```

Search completed: March 23, 2004, 14:39:23
 Job time : 47 secs

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:32:47 ; Search time 45 Seconds
(without alignments)
2643.341 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEL.....PGPMDCHRALEPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query				
No.	Score	Match Length DB	ID			Description
<hr/>						

1	1415	71.5	373	11	Q9CXB2	Q9cxb2 mus musculu
2	883	44.6	361	13	Q7T3A9	Q7t3a9 brachydanio
3	871	44.0	348	11	P70172	P70172 mus musculu
4	866	43.8	348	11	Q925U7	Q925u7 mus musculu
5	567.5	28.7	348	6	O97736	O97736 oryctolagus
6	546	27.6	317	11	O35940	O35940 mus musculu
7	535	27.0	437	4	Q8WUZ2	Q8wuz2 homo sapien
8	535	27.0	462	4	Q96EP9	Q96ep9 homo sapien
9	516	26.1	437	11	Q8BJC7	Q8bjc7 mus musculu
10	363	18.3	143	6	Q7YS68	Q7ys68 oryctolagus
11	352	17.8	125	11	Q8VI83	Q8vi83 mus musculu
12	351.5	17.8	473	11	Q8QZR2	Q8qzr2 mus musculu
13	351	17.7	187	11	Q9QZJ2	Q9qzj2 mesocricetu
14	333.5	16.9	448	4	Q9BSL2	Q9bsl2 homo sapien
15	325	16.4	321	16	O34524	O34524 bacillus su
16	321	16.2	324	16	Q8CUW3	Q8cuw3 oceanobacil
17	301.5	15.2	305	16	Q99RV2	Q99rv2 staphylococ
18	301.5	15.2	305	16	Q8NV71	Q8nv71 staphylococ
19	301.5	15.2	318	16	Q8FZ15	Q8fz15 brucella su
20	299.5	15.1	318	16	Q8YIU6	Q8yiu6 brucella me
21	299.5	15.1	323	16	Q9KEJ4	Q9kej4 bacillus ha
22	286.5	14.5	352	16	Q7WL87	Q7wl87 bordetella
23	286.5	14.5	352	16	Q7W7U7	Q7w7u7 bordetella
24	284	14.4	338	10	O81017	O81017 arabidopsis
25	279.5	14.1	315	16	Q87YA3	Q87ya3 pseudomonas
26	277	14.0	408	10	Q9ASA9	Q9asa9 oryza sativ
27	276	13.9	409	10	Q8S2V7	Q8s2v7 arabidopsis
28	270	13.6	311	17	Q8Q0R7	Q8q0r7 methanosarc
29	269.5	13.6	311	16	Q9HYX6	Q9hyx6 pseudomonas
30	268.5	13.6	335	16	Q8NR06	Q8nr06 corynebacte
31	266.5	13.5	315	16	Q9K0A9	Q9k0a9 neisseria m
32	266.5	13.5	338	17	Q8TPB0	Q8tpb0 methanosarc
33	266	13.4	455	5	Q9VXV4	Q9vxv4 drosophila
34	265.5	13.4	272	16	Q8DJU6	Q8dju6 synechococc
35	265.5	13.4	315	16	Q9JVB5	Q9jvb5 neisseria m
36	264.5	13.4	324	16	Q8NN79	Q8nn79 corynebacte
37	264	13.3	271	10	Q84WD8	Q84wd8 arabidopsis
38	257.5	13.0	297	16	Q9I375	Q9i375 pseudomonas
39	253.5	12.8	308	16	Q8E8Y1	Q8e8y1 shewanella
40	253.5	12.8	393	16	Q7U8I3	Q7u8i3 synechococc
41	252	12.7	316	16	Q8DV97	Q8dv97 streptococc
42	251.5	12.7	322	16	Q81YH5	Q81yh5 bacillus an
43	251.5	12.7	345	16	Q8FNB5	Q8fnb5 corynebacte
44	250	12.6	297	16	O67889	O67889 aquifex aeo
45	248	12.5	353	16	Q8G5B3	Q8g5b3 bifidobacte

ALIGNMENTS

RESULT 1

Q9CXB2

ID Q9CXB2 PRELIMINARY; PRT; 373 AA.
AC Q9CXB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE 8430417G17Rik protein.
 GN 8430417G17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic lung;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK018423; BAB31203.1; -.
 DR MGD; MGI:1923000; 8430417G17Rik.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR Pfam; PF01758; SBF; 1.
 SQ SEQUENCE 373 AA; 40681 MW; 0902D18506A8AC55 CRC64;

Query Match 71.5%; Score 1415; DB 11; Length 373;
 Best Local Similarity 70.3%; Pred. No. 2.9e-108;
 Matches 265; Conservative 50; Mismatches 58; Indels 4; Gaps 2;

Qy 1 MRANCSSSSACPANSSSEELPVGLEVHGNLELVFTTVVSTVMMGLLMFSLGCSVEIRKLWS 60
 | :|: :| || ||:|:| ||:| | ||:|:|:| | ||:|:| | |||| :|||
 Db 1 MSTDCAGNSTCPVNSTEEDPPVGMEGHANLKLFTVLSAVMVGLVMFSGCSVESQKLWL 60

 Qy 61 HIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVD 120
 |:||||||| ||||| ||||| | ||| |||||:| | |||||: |||||
 Db 61 HLRRPWGIAVGLLSQFGLMPLTAYLLAIGFGLKPFQAI AVLMMGSCPGGTISNVLTFWVD 120

 Qy 121 GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQONLTIPYQNIGITLVCLTIPVAFGV 180
 ||||| ||||| ||||| |||||:| ||:| ||| ||||:|||| | :|| ||
 Db 121 GDMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQONLTIPYQNIGITLVSLVVPVAGSV 180

 Qy 181 YVNYRWPQSKIIILKIGAVVGGVLLL VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVT 240
 ||||| ||||: ||||:|:|:| ||||| |:|||| ||:|:| || ||||:||||

Db 181 YVNYRWPQKQATVILKVGAILGGMLLLVAVTGMVLAKG-WNTDVTLLVISCIFPLVGHVT 239

QY 241 GFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLID 300
 ||||| |||||:|||||:| |||||:|:|:|:| |||||:|

Db 240 GFLAFLTHQSWQRCRTISLETGAQNIQLCIAMLQLSFSAEYLVQLLNLFALAYGLFQVLH 299

QY 301 GFLIVAAAYQTYKRRLKNKHGKKNKSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGP 360
 | ||||| |||| |:: | :||: :: ||:|: | | |

Db 300 GLLIVAAAYQAYKRRQKSKCRRQHPDCPDVCYEKQ---PRETSAFLDKGDEAAVTLGPVQP 356

QY 361 MDCHRALEPVGHITSCE 377
 ||| | || |||

Db 357 EQHHRAAELTSHIPSCE 373

RESULT 2

Q7T3A9

ID Q7T3A9 PRELIMINARY; PRT; 361 AA.

AC Q7T3A9;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RA Strausberg R.;

RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.

DR EMBL; BC053189; AAH53189.1; -.
KW Hypothetical protein.
SQ SEQUENCE 361 AA; 39284 MW; 5729C89AEBABC323 CRC64;

Query Match 44.6%; Score 883; DB 13; Length 361;
Best Local Similarity 48.5%; Pred. No. 1.5e-64;
Matches 172; Conservative 62; Mismatches 107; Indels 14; Gaps 5;

```
Qy      5 C S S S S A C P A N S S E E L P V G L E ----- V H G N L E L V F T V V S T V M M G L L M F S L G C S V E I R K 57
      | :   | | | :   : |           : | : | : | | | : : | | | : | | | |
Db      2 C T L E P V C P V N A T --- I C T G T S C L V P R D P F N D I L S V V M S V A I T V M L A M V M F S M G C T V E A R K 58

Qy     158 L W S H I R R P W G I A V G L L C Q F G L M P F T A Y L L A I S F S L K P V Q A I A V L I M G C C P G G T I S N I F T F 117
      | | | : | | | | | | : | | | | | : | | | | : | | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db     159 L W G H V R R P W G I F I G F L C Q F G I M P F T A F I L S L L F N V L P V Q A V V I I M G C C P G G S S S N V F C Y 118

Qy     118 W V D G D M D L S I S M T T C S T V A A L G M M P L C I Y L Y T W S W S L Q Q N L T I P Y Q N I G I T L V C L T I P V A 177
      | : | | | | | | | | | | | | : | | | | | : | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db     119 W L D G D M D L S I S M T A C S S I L A L G M M P L C L L I Y T T I W T A G D A I Q I P Y D N I G I T L V S L L V P V G 178

Qy     178 F G V Y V N Y R W P K Q S K I I L K I G A V V G G V L L L V V A V A G V V L A K G S W N S D I T L L T I S F I F P L I G 237
      | : | : | | | : | | | | : | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db     179 L G M L V K H K W P K A A K I L K V G S V V G I V L I I V I A V I G G V L Y Q S S W T I A P S L W I I G T I Y P F I G 238

Qy     238 H V T G F L L A L F T H Q S W Q R C R T I S L E T G A Q N I Q M C I T M L Q L S F T A E H L V Q M L S F P L A Y G L F Q 297
      | | | | | | | | | | | | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | |
Db     239 F G L G F L L A R F V G Q P W H R C R T I A L E T G M Q N A Q L A S T I T Q L S F S P A E L E V M F A F P L I Y S I F Q 298

Qy     298 L I D G F L I V A A Y Q T Y K R - R L K N K H G K N S G C T E V C H T R K S T S S R E T N A F L E V N E E G 351
      | :   : | : : : | | | :   : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db     299 L V V A G I A V S I H Y S I K R C R H Q T L V E E D G E G T T E D C D -- K H S Y S L E N G G F - S C D E N G 350
```

RESULT 3

P70172

ID P70172 PRELIMINARY; PRT; 348 AA.
AC P70172;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ileal NA+-dependent bile acid transporter (ISBT).
GN SLC10A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RA Saeki T., Matoba K.;
RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 9-332 FROM N.A.
RC STRAIN=ICR;
RA Saeki T.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB002693; BAA19606.1; -.

DR EMBL; D87059; BAA13237.1; -.
 DR MGD; MGI:1201406; Slc10a2.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0015711; P:organic anion transport; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR InterPro; IPR004710; Bil_ac_transpt.
 DR Pfam; PF01758; SBF; 1.
 DR TIGRFAMs; TIGR00841; bass; 1.
 SQ SEQUENCE 348 AA; 38134 MW; D00B5E43431875D7 CRC64;

Query Match 44.0%; Score 871; DB 11; Length 348;
 Best Local Similarity 47.4%; Pred. No. 1.4e-63;
 Matches 167; Conservative 74; Mismatches 97; Indels 14; Gaps 5;

Qy 7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
 :|| || |:: | : | | : | | : | |:: ::|||:|:|:| | ||
 Db 3 NSSVCPNATVCEGDSCVVPESNFNAILNTVMSTVLTILLAMVMFSMGCNVEVHKFLGHI 62
 Qy 63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIETFWVDGD 122
 :|||| | | ||||:| | :|::: : ||||: ||||| |||| | | :|:| |
 Db 63 KRPWGI FVGF L C Q F G I M P L T G F I L S V A S G I L P V Q A V V L I M G C C P G G T G S N I L A Y W I D G D 122
 Qy 123 MDLSISM TTCSTVAALGMMPLCIYLTWSWSLQQNL TIPYQNIGITLVCLTIPVAFGVYV 182
 ||||:|||||: |||||:|:| | : ||| :||:| | | ||:|:|:|
 Db 123 MDLSVSM TTCSTLLALGMMPLCLFVYTKMWVDSGTIVIPYDSIGISLVALVIPVSFGMFV 182
 Qy 183 NYRWPQKSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
 |:|:|:|:|:|:|:|:|:| :|:|:|:| | :| : :| : | | ||:| : ||
 Db 183 NHKWPQKAKIILKIGSITGVILIVLIAVIGGILYQSAWIIIEPKLWIIGTIFPIAGYSLGF 242
 Qy 243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLS FPLAYGLFQLIDGF 302
 || | | ||||:| || | |:| |:|:|:| : | | : :|| | :||:|
 Db 243 FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA 302
 Qy 303 LIVAAYQTYKRRLKNKHGKKN SGCTEVCHTRKSTSSR----ETNAFLEVNEE 350
 :|: | |:| : :|| : : | | || || : :|
 Db 303 VILGIYV TYRK----CYGKNDAEFLE--KTDNEMDSRPSFDETNKGFQ PDEK 348

RESULT 4

Q925U7

ID Q925U7 PRELIMINARY; PRT; 348 AA.
 AC Q925U7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ileal sodium-dependent bile acid transporter.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL;
 RA Dawson P.A., Craddock A.L., Tietjen M.E., Haywood J.H.;

```

RT      "Disruption of the Ileal Bile Acid Transporter Gene in Mice.";
RL      Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AF271073; AAK55514.1; -.
DR      EMBL; AF266724; AAK55514.1; JOINED.
DR      EMBL; AF266725; AAK55514.1; JOINED.
DR      EMBL; AF266726; AAK55514.1; JOINED.
DR      EMBL; AF266727; AAK55514.1; JOINED.
DR      EMBL; AF266728; AAK55514.1; JOINED.
DR      GO; GO:0016021; C:integral to membrane; IEA.
DR      GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
DR      GO; GO:0015711; P:organic anion transport; IEA.
DR      GO; GO:0006814; P:sodium ion transport; IEA.
DR      InterPro; IPR002657; BilAc/Na_symport.
DR      InterPro; IPR004710; Bil_ac_transpt.
DR      Pfam; PF01758; SBF; 1.
DR      TIGRFAMs; TIGR00841; bass; 1.
SQ      SEQUENCE      348 AA;  38094 MW;  AD33A1BC76A44482 CRC64;

```

Qy	7	SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMSFLGCSVEIRKLWSHI	62
Db	3	NSSVCPPNATVCEGDSCVVPESNFENAVLNTVMSSVLTTLLAMVMFMSGCNVEVHKFLGHI	62
Qy	63	RRPWGIAVGLLCCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD	122
Db	63	KRPWGIFVGFLLCCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD	122
Qy	123	MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPIYQNI GITLVCLTIPVAFGVYV	182
Db	123	MDLSVSMTTCSTLLALGMMPLCLFVYTKMWVDSGTIVIPYDSIGISLVALVIPVSFGMFV	182
Qy	183	NYRWPQKQSKIILKIGAVVGGVLLLVVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF	242
Db	183	NHKWPQKAKIILKIGSITGVILIVLIAVIGGILYQSAWIIIEPKLWIIGTIFPIAGYSLGF	242
Qy	243	LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGF	302
Db	243	FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA	302
Qy	303	LIVAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSR----ETNAFLEVNEE	350
Db	303	VILGIYVTVYRK----CYGKNDAEFLE--KTDNEMDSRPSFDETNGKFOPDEK	348

097736

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stengelin S., Becker W., Maier M., Noll R., Kramer W.;
 RT "Rabbit cDNA encoding hepatic sodium-dependent bile acid
 RT transporter.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AJ131361; CAA10360.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0015711; P:organic anion transport; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR InterPro; IPR004710; Bil_ac_transpt.
 DR Pfam; PF01758; SBF; 1.
 DR TIGRFAMs; TIGR00841; bass; 1.
 SQ SEQUENCE 348 AA; 37932 MW; 992A08F4AAA4489B CRC64;

Query Match 28.7%; Score 567.5; DB 6; Length 348;
 Best Local Similarity 38.9%; Pred. No. 1.2e-38;
 Matches 112; Conservative 69; Mismatches 100; Indels 7; Gaps 3;

Qy 31 ELVFTTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISE 90
 :| :|: :|: :| |||:| | :| :| :|:| :| :|:| ||:| |
 Db 24 DLALSVILVIMLLTIMLSLGCTMEFSKIKAHFLKPKGLAIALVAQYGIMPLTAFVLGKVF 83

 Qy 91 SLKPVQAIHAVLINGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYT- 149
 : :|:|:|:| || ||| :|:|:| | |||:| || |||| | ||||| :|:|:
 Db 84 RMNNIEALAILVCGCSPGGNMSNLFSLAVKGDMLNSIVMTTCSTFLALGMMPLLLYIYSR 143

 Qy 150 --WSWSLQQNLTIPTYQNIGITLVCLTIPVAFGVVNYRWPKQSKIILKIGAVVGGVLLLV 207
 : | : :|| | | :| : || | : : :| : :|:| : : : :
 Db 144 GIYEGDLKDK--VPYGGIMISIVMILIPCTIGIILKSKRPQYVPYVIKVGIIITFSISIA 201

 Qy 208 VAVAGVVLAKGSWNSDIT--LLTISFIFPLIGHVTGFELLALFTHQSWQRCRTISLETGAQ 265
 ||| : | :| |||| : | || : ||:| : | : |||:| || |
 Db 202 VAVLSAINVGNSIMYVMTPTYFLTISALMPFIFGLLGFI LSALFRLSARCSRTISMETGFQ 261

 Qy 266 NIQMCITMLQLSFTAHLVQMLSFPPLAYGLFQLIDGFLIVAAYQTYKR 313
 |:|:| |:| :| | : : ||| | :||| :| ||:| : :|:|
 Db 262 NVQLCSTILNVTFAPFVIGPLFFFLLYMIFQLAEGLLIIAVFRCEYK 309

RESULT 6

O35940

ID O35940 PRELIMINARY; PRT; 317 AA.
 AC O35940;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Na/taurocholate cotransporting polypeptide 2.
 GN SLC10A1 OR NTCP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC012048; AAH12048.1; -.
 DR Genew; HGNC:22980; SLC10A4.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR Pfam; PF01758; SBF; 1.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SQ SEQUENCE 462 AA; 49035 MW; B916D68AEE40622C CRC64;

Query Match 27.0%; Score 535; DB 4; Length 462;
 Best Local Similarity 35.7%; Pred. No. 7.7e-36;
 Matches 111; Conservative 59; Mismatches 101; Indels 40; Gaps 5;

QY 27 HGNLELVFTTVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLL 86
 || | : |:| |||::: |:||| | : |||||:| |:||
 Db 128 HGLNVFVGAALCITMLG-----LGCTVDVNHFGAHVRRPVGALLAALCQFGLLPLLAFL 182
 QY 87 AISFSLKPVQAI AVLIMGCCPGGTISNIFTFVWDGMDLSISMTTCSTVAALGMMPLCIY 146
 |:| | | |:|||: ||||| :||: : |||||:| | ||: || :|||::
 Db 183 ALAFKLDEVA AVALLCGCCPGGNLSNLMSLLVDGDMNLSIIMTISSTLLALVIMPLCLW 242
 QY 147 LYTWSW---SLQQNLTIPTQYQNGITLVCLTIPVAFGVVNYRWPQSKIILKI----- 196
 :|:|:| : | :| : :|| ||: ||:: |: : : |:|:
 Db 243 IYSWAWINTPIVQ--LLPLGTVTLTLCSTLPIGLGVFIRYKYSRVADYIVKVSLWSLLV 300
 QY 197 -----GAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLLALF 247
 |::| || : | |:| || |: :|: ||
 Db 301 TLVVLFIIMTGTMLGPPELLASIPA AVYVIA-----IFMPLAGYASGYGLATL 346
 QY 248 THQSWQRCRTISLETGAQNIQMCITMLQLSFTAHLVQMLSFPLAYGLFQLIDGFLIVAA 307
 | ||: |||:|:|:| :|:|:| : : | ||| | ||| : : |
 Db 347 FHLPPNCKRTVCLETGSQNVQLCTAILKLAFPPQFIGSMYMFLLYALFQSAEAGIFVLI 406
 QY 308 YQTYKRRLKNK 318
 |:| : :|
 Db 407 YKMYGSEMLHK 417

RESULT 9

Q8BJC7

ID Q8BJC7 PRELIMINARY; PRT; 437 AA.

AC Q8BJC7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Hypothetical sodium bile acid symporter containing protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ileum;
 RA Weihrauch D., Ao M., Rao M.C.;
 RT "Expression of bile acid transporter and bile acid receptors in the
 RT developing rabbit colon."
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY292653; AAP49247.1; -.
 SQ SEQUENCE 143 AA; 15339 MW; 037ED60C2E3CAF5F CRC64;

Query Match 18.3%; Score 363; DB 6; Length 143;
 Best Local Similarity 52.7%; Pred. No. 3.2e-22;
 Matches 68; Conservative 24; Mismatches 31; Indels 6; Gaps 3;

Qy 11 CPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPW 66
 | ||:: | | : | | : | | :: | | :: | | :: | | :: | | |||||
 Db 8 CLANATVCEGASCVAPESNFNAILS SVVLSTVLTILLALVMFSMGCNVEIKKFLGHIRRPW 67
 Qy 67 GIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLS 126
 || : | ||||: || | : ||:: | : | :: | ||||| |||| | | : |||||
 Db 68 GIFIGFLCQFGIMPLTGFVLAVAFGIMPIQAVVVLIMGCCPGGTASNILAYWVDGMDLR 127
 Qy 127 --ISMTTCS 133
 : : ||
 Db 128 YFLGCSCCS 136

RESULT 11

Q8VI83

ID Q8VI83 PRELIMINARY; PRT; 125 AA.
 AC Q8VI83;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ileal sodium-dependent bile acid transporter (Fragment).
 GN ISBT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Saeki T., Kirifuji K., Kanamoto R., Iwami K.;
 RT "Identification of transcription start sites in mouse ileal sodium-
 RT dependent bile acid transporter gene."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB078635; BAB84081.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR Pfam; PF01758; SBF; 1.
 FT NON_TER 125 125
 SQ SEQUENCE 125 AA; 13275 MW; C7F8EFC459D4C8F7 CRC64;

Query Match 17.8%; Score 352; DB 11; Length 125;
 Best Local Similarity 52.0%; Pred. No. 2.2e-21;
 Matches 64; Conservative 25; Mismatches 30; Indels 4; Gaps 2;

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
        :|| || |:: | : | | : | | : | |::: ::|||:|:|:| | ||
Db      3 NSSVCPNATVCEGDSCVVPESNFNAILNTVMSTVLTILLAMVMFSMGCNVEVHKFLGHI 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
        :||||| || |||||:| | :|:|:: : |||: ||||| ||||| ||| :|:| |
Db     63 KRPWGI FVGFLCQFGIMPLTGFILSVASGILPVQAVVVLIMGCCPGGTGSNILAYWIDGD 122

Qy    123 MDL 125
        |||
Db    123 MDL 125

```

RESULT 12

Q8QZR2

```

ID   Q8QZR2          PRELIMINARY;          PRT;    473 AA.
AC   Q8QZR2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to protein P3 (Hypothetical protein).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; BC023050; AAH23050.1; -.
DR   EMBL; BC027440; AAH27440.1; -.
DR   EMBL; AK041958; BAC31110.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
DR   GO; GO:0015711; P:organic anion transport; IEA.
DR   GO; GO:0006814; P:sodium ion transport; IEA.
DR   InterPro; IPR002657; BilAc/Na_symport.
DR   InterPro; IPR004710; Bil_ac_transpt.
DR   Pfam; PF01758; SBF; 1.
DR   TIGRFAMs; TIGR00841; bass; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 473 AA; 50254 MW; 9A2AD0A005DD1805 CRC64;

```

Query Match 17.8%; Score 351.5; DB 11; Length 473;
 Best Local Similarity 32.4%; Pred. No. 9.5e-21;
 Matches 91; Conservative 57; Mismatches 108; Indels 25; Gaps 5;

```

Qy      11 CPANSSEELPVGLEVH-GNLE---LVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPW 66
      |  |  | :||  |  : | :      : : : :  |  || ||:  |  : : |
Db      163 CIRVSPAEDLPSALNTNLGHFSENPIYLLLLPLIFVNKCSF--GCKVELEVLEKELLQSPQ 220

Qy      67 GIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLS 126
      : : ||| || : ||| | : | |  |  | : : |  ||| | : | : ||: | :
Db      221 PMLLGLLGQFLVMPFYAFLMAKVFM LPKALALGLIITCSSPGGGGSYLFSLLLGGDVTLA 280

Qy     127 ISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVVNYRW 186
      |||| ||||| | : ||  : | : | : | : |  |  || : | : | || : :
Db     281 ISMTFISTVAATGFLPLSSAIYSYLLSIHETLHVPISKILGTLLFIAIPAAAGVVIKSKL 340

Qy     187 PKQSKIILKIGAVVGGVLLL-----VVAVAGVVLAKGSWNSDITLLTISFIFPLI 236
      || | : : | : :      : |||  |  : || |  : : |  || :
Db     341 PKFSELLQVIKPF SFILLGLFLAYHMGVFILVGVR-----PIVLVGFTVPLV 391

Qy     237 GHVTGFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLS 277
      | : | : || :      : || : | : | ||  : : | |||
Db     392 GLLVGYSLAICLKL PVAQRRTVSIEVGVQNSLLALAMLQLS 432
  
```

RESULT 13

Q9QZJ2

```

ID  Q9QZJ2      PRELIMINARY;      PRT;      187 AA.
AC  Q9QZJ2;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Na-taurocholate cotransporting polypeptide (Fragment).
OS  Mesocricetus auratus (Golden hamster).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC  Mesocricetus.
OX  NCBI_TaxID=10036;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Balasubramanian N., Arrese M., Suchy F.J., Ananthanarayanan M.;
RT  "Na-Taurocholate cotransporting polypeptide (Ntcp) from Hamster
RT  liver.";
RL  Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF181258; AAD53961.1; -.
DR  GO; GO:0016020; C:membrane; IEA.
DR  GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
DR  GO; GO:0006814; P:sodium ion transport; IEA.
DR  InterPro; IPR002657; BilAc/Na_symport.
DR  Pfam; PF01758; SBF; 1.
FT  NON_TER      1      1
FT  NON_TER      187     187
SQ  SEQUENCE      187 AA;  20182 MW;  2855C5F44AB482C6 CRC64;
  
```

Query Match 17.7%; Score 351; DB 11; Length 187;
 Best Local Similarity 42.6%; Pred. No. 4.1e-21;
 Matches 80; Conservative 34; Mismatches 64; Indels 10; Gaps 5;

Qy	75	QFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWDGMDLSISMTTCST	134
		: : : : : : : :	
Db	1	QFGIMPLAAFVLGKVFHLKPIEALAILICGCSPPGNLSNFLT LAMKGDMNLSIVMTTCST	60
Qy	135	VAALGMMPLCIYLYT---WSWSLQQNLTI PYQNIGITLVCLTIPVAFGVVNYRWPKQSK	191
		: : : : : : : : : :	
Db	61	FAALGMMPLLLYIYTKGIYDGD LKDK--VPYGGIMISLVMVLIPCSLGI FLKTKRPQYVP	118
Qy	192	IILKIGAVVGGVLLLVAVAGVVLAKG SWNSDIT--LLTISFIFPLIGHVTG FLL-ALFT	248
		: : : : : : : : :	
Db	119	YIIKGGMTITFLLSVAVTVLSI INVGNSIKFAMTPPLLATSSIMPFSGFLLGYALSALF-	177
Qy	249	HQSWQRCR	256
Db	178	-QLNPRCR	184

Query Match 16.9%; Score 333.5; DB 4; Length 448;
Best Local Similarity 31.5%; Pred. No. 2.7e-19;
Matches 87; Conservative 53; Mismatches 115; Indels 21; Gaps 3;

Qy 132 CSTVAALGMMPLCIYLYTWSWSLQONLTIPYQNIGITLVCLTIPVAFGVVNYRWPKQSK 191
 |||| | :|| :|: | : | :| | ||: : ||:| || : : || |:
 Db 261 LSTVAATGFLPLSSAIYSRLLSIHETLHVPISKILGTLFLFIAPIAVGVLIKSKLPKFSQ 320

Qy 192 IILKIGAVVGGVLLL-----VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTG 241
 ::||: ||| | :||: | : : ||:| : |
 Db 321 LLLQVVKPFSFVLLLGGFLAYRMGVFILAGIRL-----PIVLVGITVPLVGLLVG 371

Qy 242 FLLALFTHQSWQRCRTISLETGAQNIQMCITMLQLS 277
 : || : ||:|:| | || : : |||||
 Db 372 YCLATCLKLPVAQRRTVSIEVGVQNSLLALAMLQLS 407

RESULT 15

O34524

ID O34524 PRELIMINARY; PRT; 321 AA.
 AC O34524;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative transporter.
 GN YOCS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lapidus A., Galleron N., Sorokin A., Ehrlich D.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF027868; AAB84443.1; -.
 DR EMBL; Z99114; CAB13827.1; -.
 DR PIR; E69902; E69902.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; IEA.
 DR GO; GO:0015711; P:organic anion transport; IEA.
 DR GO; GO:0006814; P:sodium ion transport; IEA.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR InterPro; IPR004710; Bil_ac_transpt.
 DR Pfam; PF01758; SBF; 1.
 DR TIGRFAMs; TIGR00841; bass; 1.
 KW Complete proteome.
 SQ SEQUENCE 321 AA; 34251 MW; 0D9CCF6B36E84A96 CRC64;

Query Match 16.4%; Score 325; DB 16; Length 321;
 Best Local Similarity 27.6%; Pred. No. 9.7e-19;
 Matches 84; Conservative 76; Mismatches 114; Indels 30; Gaps 12;

Qy	33	VFTVVS---TVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAIS	89
		: : : : : : : : : :	
Db	32	LFTWISSYITIFLGIIMFGMGLTLQADDFKELVRKPWQVIIGVIAQYTIMPLVAFGLAFG	91
Qy	90	FSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYT	149
		: : : : : : : :	
Db	92	LHLPAEIAVG VILVGCCPGGTASNVM TFLAKGNTALSVAVTTISTLLAPVVTPLLIMLFA	151
Qy	150	WSWSLQQNL TIPYQNIGITLV-CLTIPVAFGVVNYRWPKQ-SKII--LKIGAVVGGVLL	205
		: : : : : : : : : : : : :	
Db	152	KEW-----LPVSPGSLFISILQAVLFPIIAGLIVKMFFRKQVAKAVHALPLVSVIG----	202
Qy	206	LVVAVAGVVLAKGSWN---SDITLLTISFIFPLIGHVTGFLALFTHQSWQRCRTISLET	262
		: : : : : : : : : :	
Db	203	-IVAIVSAVVSGNRENLLQSGLLIFSVVILHNGIGYLLGFLCAKLLKMDYPSQKAIAIEV	261
Qy	263	GAQNIQMCITMLQLSFTAHLVQMLSFPLA-YGLFQLIDGFLIVAAYQTYKRRLKNKH-G	320
		: : : : : : : : : : : :	
Db	262	GMQN-----SGLGAALATAHFSPLSAVPSAIFSVVHNLSGSML-ATY--WSKKVKKKQAG	313
Qy	321	KKNS	324
		:	
Db	314	SKSS	317

Search completed: March 23, 2004, 14:37:17

Job time : 48 secs

OM protein - protein search, using sw model

Run on: March 23, 2004, 14:32:07 ; Search time 17 Seconds
(without alignments)
1154.732 Million cell updates/sec

Title: US-10-091-628-2
Perfect score: 1979
Sequence: 1 MRANCSSSSACPANSSEEL.....PGPMDCHRALEPVGHITSCE 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	886	44.8	347	1 NTCI_RABIT	Q28727 o ileal sod
2	884	44.7	348	1 NTCI_CRIGR	Q60414 c ileal sod
3	871	44.0	348	1 NTCI_RAT	Q62633 r ileal sod
4	860.5	43.5	348	1 NTCI_HUMAN	Q12908 h ileal sod
5	559.5	28.3	362	1 NTCP_RAT	P26435 rattus norv
6	553	27.9	349	1 NTCP_HUMAN	Q14973 homo sapien
7	553	27.9	362	1 NTCP_MOUSE	O08705 mus musculu
8	333.5	16.9	477	1 P3_HUMAN	P09131 homo sapien
9	180.5	9.1	182	1 P3_MOUSE	P21129 mus musculu
10	125	6.3	409	1 YCXA_BACSU	Q08791 bacillus su
11	117	5.9	721	1 YJIY_ECOLI	P39396 escherichia
12	116	5.9	286	1 YCXE_BACME	P40419 bacillus me
13	112.5	5.7	368	1 CYB_TOXGO	O20672 toxoplasma
14	109.5	5.5	383	1 Y944_SYNY3	P74311 synechocyst
15	107.5	5.4	576	1 NU5M_ANOQU	P33510 anopheles q
16	106.5	5.4	443	1 COXX_HUMAN	Q12887 homo sapien
17	104.5	5.3	397	1 YGED_ECOLI	P39196 escherichia

18	104	5.3	551	1	LLDP_ECOL6	Q8xdf9	escherichia
19	104	5.3	551	1	LLDP_ECOLI	P33231	escherichia
20	103	5.2	372	1	OPRD_HUMAN	P41143	homo sapien
21	101.5	5.1	532	1	YC18_HAEIN	Q57251	haemophilus
22	101	5.1	379	1	CYB_PONBL	Q9mg84	pontoporia
23	100.5	5.1	332	1	YFEH_ECOLI	P39836	escherichia
24	99.5	5.0	379	1	CYB_BOLAM	P21722	bolomys amo
25	99.5	5.0	387	1	GALS_HUMAN	O43603	homo sapien
26	99.5	5.0	436	1	SECY_METJA	Q60175	methanococc
27	99.5	5.0	541	1	NU5M_ARTSF	Q37710	artemia san
28	99	5.0	312	1	CCSA_ODOSI	P49523	odontella s
29	99	5.0	333	1	YK26_PSEAE	P39879	pseudomonas
30	98.5	5.0	391	1	SOTB_HELPJ	Q9zk31	helicobacte
31	98	5.0	391	1	YIDY_ECOLI	P31462	escherichia
32	98	5.0	835	1	CD97_HUMAN	P48960	homo sapien
33	97.5	4.9	460	1	YICJ_ECOLI	P31435	escherichia
34	97	4.9	369	1	MRAY_ANASP	Q8yp83	anabaena sp
35	96.5	4.9	404	1	SOTB_PASMU	Q9cm87	pasteurella
36	96	4.9	444	1	YOJA_BACSU	O31862	bacillus su
37	96	4.9	465	1	DCDR_XENLA	P42291	xenopus lae
38	96	4.9	572	1	FZD7_MOUSE	Q61090	mus musculu
39	95.5	4.8	246	1	AQPM_METTH	O26206	methanobact
40	95.5	4.8	381	1	CYB_THODA	Q9xnx1	thomasomys
41	95.5	4.8	551	1	LLDP_SALTI	Q8z2e3	salmonella
42	95.5	4.8	551	1	LLDP_SALTY	Q8zl63	salmonella
43	95	4.8	347	1	NU2M_CANFA	Q9zz65	canis famil
44	95	4.8	380	1	CYB_ONCKE	Q9xm10	oncorhynchu
45	95	4.8	380	1	CYB_ONCMY	P48173	oncorhynchu

ALIGNMENTS

RESULT 1

NTCI_RABIT

ID NTCI_RABIT STANDARD; PRT; 347 AA.
AC Q28727;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ileal sodium/bile acid cotransporter (Ileal Na(+)/bile acid
DE cotransporter) (Na+ dependent ileal bile acid transporter) (Ileal
DE sodium-dependent bile acid transporter) (ISBT) (Sodium/taurocholate
DE cotransporting polypeptide, ileal).
GN SLC10A2 OR NTCP2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Ileum;
RA Stengelin S., Apel S., Becker W., Maier M., Rosenberger J., Wess G.,
RA Kramer W.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Plays a critical role in the sodium-dependent
CC reabsorption of bile acids from the lumen of the small intestine.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
CC (SBF).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z54357; CAA91184.1; -.
DR EMBL; AJ002005; CAA05135.1; -.
DR InterPro; IPR004710; Bil_ac_transpt.
DR InterPro; IPR002657; BilAc/Na_symport.
DR Pfam; PF01758; SBF; 1.
DR TIGRFAMs; TIGR00841; bass; 1.
KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 290 310 POTENTIAL.
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 11 11 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 347 AA; 37729 MW; 1315BB6BADDEE66C CRC64;

Query Match 44.8%; Score 886; DB 1; Length 347;
Best Local Similarity 47.0%; Pred. No. 9.4e-61;
Matches 164; Conservative 73; Mismatches 98; Indels 14; Gaps 5;

Qy 11 CPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPW 66
| ||:: | | : | | : | | :: | :::: |::|::|::|::| | |||||
Db 8 CLANATVCEGASCVAPESNFNAILS SVVLSTVLTLALLVMFMSMGCNVEIKKFLGHIRRPW 67

Qy 67 GIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLS 126
|| :| ||||::|| | ::||::| : |::|: ||||| ||||| ||| :||| |||||
Db 68 GIFIGFLCQFGIMPLTGFVLAVAFGIMPIQAVVVLIMGCCPGGTASNILAYWVDGMDLS 127

Qy 127 ISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNIGITLVCLTIPVAFGVYVNYRW 186
:|||||||: |||||::|::| | : || | || :|| | :||: |::|::|
Db 128 VSMTTCSTLLALGMMPLCLYVYTKMWVDSGTIVIPYDNIGTSLVALVVPVSIGMFEVNHKW 187

Qy 187 PKQSKIILKIGAVVGGVLLL VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLAL 246
|:::| ||||:|:: | ||:::| | :| : : | | | ||| :| : || ||
Db 188 PQKAKIILKVGSIAGAVLIVLIAVVG GILYQSAWII EPKLWIIGTIFPMAGYSLGFFLAR 247

Qy 247 FTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGFLIVA 306
| | ||||:| || | |::| ||:: | | : :|| | :||: : :
Db 248 IAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLTYVFTFPLIYSIFQIAFAAIFLG 307

Qy 307 AYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETN AFLEVNEEGAITP 355
| |:: ||| :: : : | | : :|| ::| | |

RESULT 2

NTCI_CRIGR

ID NTCI_CRIGR STANDARD; PRT; 348 AA.
AC Q60414;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Ileal sodium/bile acid cotransporter (Ileal Na(+)/bile acid
DE cotransporter) (Na+ dependent ileal bile acid transporter) (Ileal
DE sodium-dependent bile acid transporter) (ISBT) (Sodium/taurocholate
DE cotransporting polypeptide, ileal).
GN SLC10A2 OR NTCP2.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ileum;
RX MEDLINE=94117449; PubMed=8288599;
RA Wong M.H., Oelkers P.M., Craddock A.L., Dawson P.A.;
RT "Expression cloning and characterization of the hamster ileal sodium-
RT dependent bile acid transporter.";
RL J. Biol. Chem. 269:1340-1347(1994).
CC -!- FUNCTION: Plays a critical role in the sodium-dependent
CC reabsorption of bile acids from the lumen of the small intestine.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
CC (SBF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U02028; AAA18640.1; -.
DR InterPro; IPR004710; Bil_ac_transpt.
DR InterPro; IPR002657; BilAc/Na_sympoort.
DR Pfam; PF01758; SBF; 1.
DR TIGRFAMs; TIGR00841; bass; 1.
KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 196 216 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 348 AA; 37919 MW; 1F3A1CFC9C8DFB8C CRC64;

Query Match 44.7%; Score 884; DB 1; Length 348;
 Best Local Similarity 46.9%; Pred. No. 1.3e-60;
 Matches 164; Conservative 74; Mismatches 102; Indels 10; Gaps 4;

```

Qy      7 SSSACPANSS--EEELPVGLEVHGN--LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
      :|| | |:: | : : | : | | : | |::: |::|::|::|: | |:
Db      3 NSSICNPNATICEGDSCIAPESNFNAILSVVMSTVLTILLALVMFMSMGCNVELHKFLGHL 62

Qy     63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSILKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
      ||||| | | |||:| | ::::| : |||: || | ||||| | | :|||
Db     63 RRPWGI VVGFLCQFGIMPLTGFVLSVAFGILPVQAVVVL IQGCCPGGTASNILAYWVDGD 122

Qy    123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNI GITLVCLTIPVAFGVYV 182
      |||:|||||: |||||:::| | : |||: || :| | |||: |::
Db    123 MDLSVSMTTCTLLALGMMPLCLFIYTKMWVDSGTIVIPYDSIGTSLVALVIPVSIGMYV 182

Qy    183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
      |::|:::|||||:: | ::::| | :| :| : | | |:: | : ||
Db    183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWTIEPKLWIIGTIYPIAGYGLGF 242

Qy    243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGF 302
      || | | |||:|||| | | :| |::|::| | | : ||| | :||
Db    243 FLARIAGQPWYRCRTVALETGLQNTQLCSTIVQLSFSPEDLNLVFTFPLIYSIFQIAFAA 302

Qy    303 LIVAAAYQTYKRRLKNKHGKKN SGCTEVCHTRKS--TSSRETNAFLEVNEE 350
      :: || ||: || | : | : ||| : :|
Db    303 ILLGAYVAYKK----CHGKNNTLQEKTDNEMEPRSSFQETNKGFPDEK 348
  
```

RESULT 3

NTCI_RAT

ID NTCI_RAT STANDARD; PRT; 348 AA.
 AC Q62633;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Ileal sodium/bile acid cotransporter (Ileal Na(+)/bile acid
 DE cotransporter) (Na+ dependent ileal bile acid transporter) (Ileal
 DE sodium-dependent bile acid transporter) (ISBT) (Sodium/taurocholate
 DE cotransporting polypeptide, ileal).
 GN SLC10A2 OR NTCP2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Ileum;
 RX MEDLINE=95164708; PubMed=7860756;
 RA Shneider B.L., Dawson P.A., Christie D.M., Hardikar W., Wong M.H.,
 RA Suchy F.J.;
 RT "Cloning and molecular characterization of the ontogeny of a rat
 RT ileal sodium-dependent bile acid transporter.";

RL J. Clin. Invest. 95:745-754(1995).
 CC -!- FUNCTION: Plays a critical role in the sodium-dependent
 CC reabsorption of bile acids from the lumen of the small intestine.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Transcriptionally regulated increases in mRNA
 CC and protein levels at the time of weaning.
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
 CC (SBF).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U07183; AAC53101.1; -.
 DR InterPro; IPR004710; Bil_ac_transpt.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR Pfam; PF01758; SBF; 1.
 DR TIGRFAMs; TIGR00841; bass; 1.
 KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.
 FT TRANSMEM 29 49 POTENTIAL.
 FT TRANSMEM 58 78 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 127 147 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 225 245 POTENTIAL.
 FT TRANSMEM 285 305 POTENTIAL.
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 348 AA; 38024 MW; D4C38CF136D1143B CRC64;

Query Match 44.0%; Score 871; DB 1; Length 348;
 Best Local Similarity 47.1%; Pred. No. 1.3e-59;
 Matches 165; Conservative 70; Mismatches 105; Indels 10; Gaps 3;

Qy 7 SSSACPANSSEELPVGLEVHGN----LELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHI 62
 :|| | |: | | | | | | : | |:: :|||:|:| | ||
 Db 3 NSSVCSPNATFCEGDSCLVTESNFNAILSTVMSTVLTILLAMVMFSMGCNVEINKFLGHI 62

 Qy 63 RRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGD 122
 :||| | | |||:| | |::: : |||: ||||| | | | :|:| |
 Db 63 KRPWGIFVGFLLCQFGIMPLTGFI LSVASGILPVQAVVVLIMGCCPGGTG SNILAYWIDGD 122

 Qy 123 MDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNL TIPYQNIGITLVCLTIPVAFGVYV 182
 |||:| | | | : | | | |::: | | : || | :|:| | | | :|:|
 Db 123 MDLSVSMTCSTLLALGMMPLCLFIYTKMVVDSGTIVIPYDSIGISLVALVIPV SIGMFV 182

 Qy 183 NYRWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGF 242
 |::|:|:| | | |:: | :|:|:| | | : : | | | | :|:|
 Db 183 NHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQSAWIIEPKLWIIGTIFPIAGYSLGF 242

 Qy 243 LLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPLAYGLFQLIDGF 302
 || | | | |::| | | | :| | |:| | | : | | :| | | :| | :

Db 243 FLARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFSPEDLNLVFTFPLIYTVFQLVFAA 302

Qy 303 LIVAAAYQTYKRRLKNKHGKKN SGCTEVCHTRKS--TSSRETNAFLEVNEE 350
 :|: | ||: ||: | :||| : :|:

Db 303 IILGMVVTYKK----CHGKNDAEFLEKTDNDMDPMPSPQETNKGFPQDEK 348

RESULT 4

NTCI_HUMAN

ID NTCI_HUMAN STANDARD; PRT; 348 AA.

AC Q12908; Q13839;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ileal sodium/bile acid cotransporter (Ileal Na+)/bile acid cotransporter) (Na+ dependent ileal bile acid transporter) (Ileal sodium-dependent bile acid transporter) (ISBT) (Sodium/taurocholate cotransporting polypeptide, ileal).

GN SLC10A2 OR NTCP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND VARIANT CD SER-290.

RC TISSUE=Ileum;

RX MEDLINE=96070831; PubMed=7592981;

RA Wong M.H., Oelkers P., Dawson P.A.;

RT "Identification of a mutation in the ileal sodium-dependent bile acid transporter gene that abolishes transport activity.";

RL J. Biol. Chem. 270:27228-27234(1995).

RN [2]

RP SEQUENCE FROM N.A., VARIANTS PBAM PRO-243 AND MET-262, AND VARIANT SER-171.

RX MEDLINE=97263517; PubMed=9109432;

RA Oelkers P., Kirby L.C., Heubi J.E., Dawson P.A.;

RT "Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2).";

RL J. Clin. Invest. 99:1880-1887(1997).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=22269935; PubMed=12364586;

RA Chumakov I., Blumenfeld M., Guerassimenko O., Cavarec L., Palicio M., Abderrahim H., Bougueleret L., Barry C., Tanaka H., La Rosa P., Puech A., Tahri N., Cohen-Akenine A., Delabrosse S., Lissarrague S., Picard F.-P., Maurice K., Essioux L., Millasseau P., Grel P., Debailleul V., Simon A.-M., Caterina D., Dufaure I., Malekzadeh K., Belova M., Luan J.-J., Bouillot M., Sambucy J.-L., Primas G., Saumier M., Boubkiri N., Martin-Saumier S., Nasroune M., Peixoto H., Delaye A., Pinchot V., Bastucci M., Guillou S., Chevillon M., Sainz-Fuertes R., Meguenni S., Aurich-Costa J., Cherif D., Gimalac A., Van Duijn C., Gauvreau D., Ouelette G., Fortier I., Realson J., Sherbatich T., Riazanskay N., Rogaev E., Raeymaekers P., Aerssens J., Konings F., Luyten W., Macciardi F., Sham P.C., Straub R.E., Weinberger D.R., Cohen N., Cohen D.;

RT "Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia.";


```

RL  Proc. Natl. Acad. Sci. U.S.A. 99:13675-13680(2002).
RN  [4]
RP  SEQUENCE FROM N.A.
RA  Smith M.;
RL  Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN  [5]
RP  SEQUENCE OF 1-22 FROM N.A.
RC  TISSUE=Blood;
RA  Stengelin S., Apel S., Becker W., Maier M., Rosenberger J.,
RA  Kaufmann C., Wess G., Kramer W.;
RL  Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC  -!- FUNCTION: Plays a critical role in the sodium-dependent
CC      reabsorption of bile acids from the lumen of the small intestine.
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC  -!- DISEASE: Defects in SLC10A2 are a cause of primary bile acid
CC      malabsorption (PBAM), an idiopathic intestinal disorder associated
CC      with congenital diarrhea, steatorrhea, interruption of the
CC      enterohepatic circulation of bile acids, and reduced plasma
CC      cholesterol levels.
CC  -!- DISEASE: Defects in SLC10A2 are a cause of Crohn's disease (CD).
CC  -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
CC      (SBF).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U10417; AAC51870.1; -.
DR  EMBL; U67674; AAC95398.1; -.
DR  EMBL; U67669; AAC95398.1; JOINED.
DR  EMBL; U67670; AAC95398.1; JOINED.
DR  EMBL; U67671; AAC95398.1; JOINED.
DR  EMBL; U67672; AAC95398.1; JOINED.
DR  EMBL; U67673; AAC95398.1; JOINED.
DR  EMBL; AE014304; AAN16026.1; -.
DR  EMBL; AL161771; CAC39447.1; -.
DR  EMBL; Z54350; CAA91161.1; -.
DR  PIR; I38655; I38655.
DR  Genew; HGNC:10906; SLC10A2.
DR  MIM; 601295; -.
DR  GO; GO:0005887; C:integral to plasma membrane; TAS.
DR  GO; GO:0008508; F:bile acid:sodium symporter activity; TAS.
DR  GO; GO:0006810; P:transport; TAS.
DR  InterPro; IPR004710; Bil_ac_transpt.
DR  InterPro; IPR002657; BilAc/Na_symport.
DR  Pfam; PF01758; SBF; 1.
DR  TIGRFAMs; TIGR00841; bass; 1.
KW  Transmembrane; Transport; Symport; Sodium transport; Glycoprotein;
KW  Disease mutation; Polymorphism.
FT  TRANSMEM      29      49      POTENTIAL.
FT  TRANSMEM      58      78      POTENTIAL.
FT  TRANSMEM      82     102      POTENTIAL.
FT  TRANSMEM     127     147      POTENTIAL.

```

FT	TRANSMEM	158	178	POTENTIAL.
FT	TRANSMEM	196	216	POTENTIAL.
FT	TRANSMEM	225	245	POTENTIAL.
FT	TRANSMEM	285	305	POTENTIAL.
FT	CARBOHYD	10	10	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	328	328	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	171	171	A -> S.
FT				/FTid=VAR_004613.
FT	VARIANT	243	243	L -> P (IN PBAM; ABOLISHES
FT				TAUROCHOLATE TRANSPORT).
FT				/FTid=VAR_004614.
FT	VARIANT	262	262	T -> M (IN PBAM; ABOLISHES
FT				TAUROCHOLATE TRANSPORT).
FT				/FTid=VAR_004615.
FT	VARIANT	290	290	P -> S (IN CD; ABOLISHES TAUROCHOLATE
FT				TRANSPORT).
FT				/FTid=VAR_004616.
SQ	SEQUENCE	348 AA;	37697 MW;	15990AAA91CCDB06 CRC64;

Qy	5	CSSSSACPANSSSEELPVGLEVHGNLELVFTVSTVMMGLLMFSLGCSVEIRKLWSHIRR	64
Db	14	CSGASCVVPE\$NFNNI-----LSVVLSTVLTILLALVMFSMGCNVEIKKFLGHIKR	64
Qy	65	PWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMD	124
Db	65	PWGICVGFLCQFGIMPLTGFILSVAFDILPLQAVVLIIGCCPGGTASNILAYWVDGDMD	124
Qy	125	LSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPIYQONIGITLVCLTIPVAFGVVNY	184
Db	125	LSVSMTTCSTLLALGMMPLCLLIYTKMWVDSGSIVIPIYDNIGTSLVALVVPV\$IGMFVNH	184
Qy	185	RWPKQSKIILKIGAVVGGVLLLVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLL	244
Db	185	KWPQKAKIILKIGSIAGAILIVLIAVVG\$ILYQSAWIIAPKLWIIIGTIFPVAGYSLGFLL	244
Qy	245	ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA\$EHLVQMLSFPLAYGLFQLIDGFLI	304
Db	245	ARIAGLPWYRCRTVAFETGMQNTQLCSTIVQLSFTPEELNVVFTFPLIYSIFQLAFAAIF	304
Qy	305	VAA\$YQTYKRRLLKNKHGKKN\$SGCTEVCHTRKSTSSRETNAFLEVNEEGAITP	355
Db	305	LGFYVAYKK----CHGKNKAEIPE----SKENGTEPESSFYKAN--GGFQP	345

RESULT 5

DE cotransporting polypeptide).
GN SLC10A1 OR NTCP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92073340; PubMed=1961729;
RA Hagenbuch B., Stieger B., Foguet M., Luebbert H., Meier P.J.;
RT "Functional expression cloning and characterization of the hepatocyte
RT Na+/bile acid cotransport system.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10629-10633(1991).
CC -!- FUNCTION: The hepatic sodium/bile acid uptake system exhibits
CC broad substrate specificity and transports various nonbile acid
CC organic compounds as well. It is strictly dependent on the
CC extracellular presence of sodium.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Liver and kidney.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
CC (SBF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77479; AAA42112.1; -.
DR PIR; A41601; A41601.
DR InterPro; IPR004710; Bil_ac_transpt.
DR InterPro; IPR002657; BilAc/Na_symport.
DR Pfam; PF01758; SBF; 1.
DR TIGRFAMs; TIGR00841; bass; 1.
KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.
FT TRANSMEM 24 45 POTENTIAL.
FT TRANSMEM 60 80 POTENTIAL.
FT TRANSMEM 82 98 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 190 211 POTENTIAL.
FT TRANSMEM 228 244 POTENTIAL.
FT TRANSMEM 285 306 POTENTIAL.
FT CARBOHYD 5 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 11 11 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 362 AA; 39295 MW; F0ABB76076A57550 CRC64;

Query Match 28.3%; Score 559.5; DB 1; Length 362;
Best Local Similarity 37.2%; Pred. No. 9.3e-36;
Matches 133; Conservative 69; Mismatches 135; Indels 21; Gaps 9;

Qy 10 ACPANSSEELPVGLEVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIA 69

```

      : | | |   | | |   |   :   : :   : | : | | | | : |   | : : | : | :
Db      7 SAPFNFS---LPPGFG-HRATDKALSIIILVLMLLIMLSLGCTMEFSKIKAHLWKPKGVI 62

Qy      70 VGLLCQFGLMPFTAYLLAISFSILKPVQAI AVLIMGCCPGGTISNIFTFWVDGMDLSISM 129
      | | : | | | : | |   | |   : : | : | | | | | | : | | : | | | |
Db      63 VALVAQFGIMPLAAFLLGKIFHLSNIEALAILICGCSPPGGLSNLFTLAMKGDMLNLSIVM 122

Qy      130 TTCSTVAALGMMPLCIYLYT---WSWSLQQNLTIPTQYQNI GITLVCLTIPVAFGVVNYRW 186
      | | | : : | | | | | : | :   :   | :   : | : | | : | |   | : : :
Db      123 TTCSSFSALGMMPLLLYVYSKGIYDGDLDKDK--VPYKGIMISLVIVLIPCTIGIVLKSQR 180

Qy      187 PKQSKIILKIGAVVGGVLLLVAVAGVVLAKGWSNSDIT--LLTISFIFPLIGHVTGFLN 244
      |   | | | | : : | : |   | :   |   | | | : | | | : | : |
Db      181 PHYVPYILKGGMIITFLLSVAVTALSVINVGNSIMFVMTPHLLATSSLMPFSGFLMGYIL 240

Qy      245 -ALFTHQSWQRC-RTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGF 302
      | | | |   | | | | : | | | | | : | : | | : | : | | | : |
Db      241 SALF--QLNPSCRRTISMETGFQNIQLCSTILNVTFPPEVIGPLFFFPLLYMIFQLAEG 298

Qy      303 LIVAAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVN EEGAITPGPPGP 360
      | | : : | : |   |   | : : : |   | :   | | | | |
Db      299 LIIIIIFRCYEKI-----KPPKDQTKITYKAAATEDATPAALEKGTHNGNIPPLQPGP 350

```

RESULT 6

NTCP_HUMAN

ID NTCP_HUMAN STANDARD; PRT; 349 AA.

AC Q14973;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Sodium/bile acid cotransporter (Na+)/bile acid cotransporter)

DE (Na+)/taurocholate transport protein) (Sodium/taurocholate

DE cotransporting polypeptide).

GN SLC10A1 OR NTCP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=94179485; PubMed=8132774;

RA Hagenbuch B., Meier P.J.;

RT "Molecular cloning, chromosomal localization, and functional

RT characterization of a human liver Na+/bile acid cotransporter.;"

RL J. Clin. Invest. 93:1326-1331(1994).

CC -!- FUNCTION: The hepatic sodium/bile acid uptake system exhibits
 CC broad substrate specificity and transports various nonbile acid
 CC organic compounds as well. It is strictly dependent on the
 CC extracellular presence of sodium.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
 CC (SBF).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

Db 315 DK-----TKMIYTAATT-----EETIPGALGNGTYKGEDC 344

RESULT 7

NTCP_MOUSE

ID NTCP_MOUSE STANDARD; PRT; 362 AA.
AC O08705;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sodium/bile acid cotransporter (Na+)/bile acid cotransporter)
DE (Na+)/taurocholate transport protein) (Sodium/taurocholate
DE cotransporting polypeptide).
GN SLC10A1 OR NTCP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Saeki T.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Hagenbuch B.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: The hepatic sodium/bile acid uptake system exhibits
CC broad substrate specificity and transports various nonbile acid
CC organic compounds as well. It is strictly dependent on the
CC extracellular presence of sodium.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
CC (SBF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003303; BAA19846.1; -.
DR EMBL; U95131; AAB81023.1; -.
DR MGD; MGI:97379; Slc10a1.
DR InterPro; IPR004710; Bil_ac_transpt.
DR InterPro; IPR002657; BilAc/Na_symport.
DR Pfam; PF01758; SBF; 1.
DR TIGRFAMs; TIGR00841; bass; 1.
KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.
FT TRANSMEM 24 45 POTENTIAL.
FT TRANSMEM 60 80 POTENTIAL.
FT TRANSMEM 82 98 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.

FT	TRANSMEM	190	211	POTENTIAL.
FT	TRANSMEM	228	244	POTENTIAL.
FT	TRANSMEM	285	306	POTENTIAL.
FT	CARBOHYD	5	5	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	11	11	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	103	103	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	117	117	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	362 AA;	39413 MW;	7A70493E1804280F CRC64;

Qy	10	ACPANSSEELVPVGLVHGNLELVFTVVSTVMMGLLMFSLGCSVEIRKLSHRRPWGIA	69
		: : : : : : :	
Db	7	SAPFNFS---LPPGFG-HRATDTALSVILVVMLLIMLSLGCTMEFSKIKAHFWKPKGVI	62
Qy	70	VGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFWVDGDMDSLISM	129
		: : : : : : : : : : :	
Db	63	IAIVAQYGIMPLSAFLLGKVFHLTSIEALAILICGCSPPGNNLSNLF TLAMKGD MNLSIVM	122
Qy	130	TTCSTVAALGMMPLCIYLYT---WSWSLQQNLTIPIYQNIGITLVCLTIPVAFGVYVNYRW	186
		: : : : : : : : : : :	
Db	123	TTCSSFTALGMMPLLLYIYSKGIYDGDLDKDK--VPYKGIMLSLVMVLIPCAIGIFLKS KR	180
Qy	187	PKQSKIILKIGAVVGGVLLLVVAVAGVVLAKGSWNSDIT--LLTISFIFPLIGHVTGFL L	244
		: : : : : : : : : : : :	
Db	181	PHYVPYVLKAGMIITFSLSVAVTVLSVINVGNSIMFVMTPHLLATSSIMPFTGFLMGYIL	240
Qy	245	ALFTHQSWQRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYGLFQLIDGFLI	304
		: : : : : : : : : : :	
Db	241	SALFRLNPSCRRTISMETGFQNVQLCSTILNVTFPPEVIGPLFFFPLLYMIFQLAEGLLF	300
Qy	305	VAAYQTYKRRLKNKHGKKNSGCTEVCHTRKSTSSRETNAFLEVNEEGAITPGPPGPM DCH	364
		: : : : : : : : :	
Db	301	IIIFRCY---LKIKPQKDQ---TKITYKAAATEDATPAALEKGTHNGNNPPTQPG-----	349
Qy	365	RALEPVG	371
Db	350	--LSPNG	354

RESULT 8

```

ID      P3_HUMAN          STANDARD;          PRT;      477 AA.
AC      P09131;
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-MAR-1989 (Rel. 10, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      P3 protein.
GN      SLC10A3 OR P3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]

```

RP SEQUENCE FROM N.A.
 RX MEDLINE=89041548; PubMed=3186440;
 RA Alcalay M., Toniolo D.;
 RT "CpG islands of the X chromosome are gene associated.";
 RL Nucleic Acids Res. 16:9527-9556(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96311563; PubMed=8733135;
 RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
 RA Zuo L., Heiner C., Burough F.W., Ripetto M., Schlessinger D.,
 RA D'Urso M.;
 RT "Long-range sequence analysis in Xq28: thirteen known and six
 RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
 RT G6PD loci.";
 RL Hum. Mol. Genet. 5:659-668(1996).
 CC -!- FUNCTION: The ubiquitous expression and the conservation of the
 CC sequence in distant animal species suggest that the gene codes for
 CC a protein with housekeeping functions.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: TO P3 PROTEIN OF ANIMALS AND YEASTS.
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
 CC (SBF).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X12458; CAA30998.1; -.
 DR EMBL; L44140; AAA92651.1; -.
 DR PIR; S01696; S01696.
 DR Genew; HGNC:22979; SLC10A3.
 DR MIM; 312090; -.
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0008508; F:bile acid:sodium symporter activity; NAS.
 DR GO; GO:0006814; P:sodium ion transport; NAS.
 DR InterPro; IPR004710; Bil_ac_transpt.
 DR InterPro; IPR002657; BilAc/Na_symport.
 DR Pfam; PF01758; SBF; 1.
 DR TIGRFAMs; TIGR00841; bass; 1.
 KW Transmembrane; Transport; Symport.
 SQ SEQUENCE 477 AA; 50332 MW; 49CB363EB3B66A1D CRC64;

Query Match 16.9%; Score 333.5; DB 1; Length 477;
 Best Local Similarity 31.5%; Pred. No. 2.4e-18;
 Matches 87; Conservative 53; Mismatches 115; Indels 21; Gaps 3;

Qy 12 PANSSEELPVGLEVHGNLELVETVSTVMMGLLMFSLGCSVEIRKLWSHIRRPWGIAVG 71
 || : | | : :: : : | ||| : | : : |
 Db 172 PAEDTPATLSADLAHFSENPIYLLPLIFVNKCSF--GCKVELEVLLKGLMQSPQPMLLG 229
 Qy 72 LLCQFGLMPFTAYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTFVWDGDMDLISMTT 131
 || || : || | : | : | | : : | || | : | : : || : | : |||
 Db 230 LLGQFLVMPLYAF LMAKVFMLPKALALGLIITCSSPGGGGSYLFSLLLGGDVT LAISM TF 289

Qy 132 CSTVAALGMMPLCIYLYTWSWSLQQNLTIPTQYQNIQITLVCLTIPVAFGVYVNYRWPQSK 191
 |||| | :|| :|| :| :| | || : ||:| || : : || |:
 Db 290 LSTVAATGFLPLSSAIYSRLLSIHETLHVPISKILGTLFLIAPIAVGVLIKSKLPKFSQ 349

Qy 192 IILKIGAVVGGVLLL-----VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTG 241
 ::||: ||| | :||: | : : ||:| : |
 Db 350 LLLQVVKPFSFVLLLGGFLFLAYRMGVFILAGIRL-----PIVLVGITVPLVGLLVG 400

Qy 242 FLLALFTHQSWQRCRTISLETGAQNIQMCITMLQLS 277
 : || : ||:|:| | || : : |||||
 Db 401 YCLATCLKLPVAQRRTVSIEVGVQNSLLALAMLQLS 436

RESULT 9

P3_MOUSE

ID P3_MOUSE STANDARD; PRT; 182 AA.
 AC P21129;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE P3 protein (Fragment).
 GN SLC10A3 OR P3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=90307023; PubMed=1973144;
 RA Filippi M., Tribioli C., Toniolo D.;
 RT "Linkage and sequence conservation of the X-linked genes DXS253E (P3)
 RT and DXS254E (GdX) in mouse and man."
 RL Genomics 7:453-457(1990).
 CC -!- FUNCTION: The ubiquitous expression and the conservation of the
 CC sequence in distant animal species suggest that the gene codes for
 CC a protein with housekeeping functions.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY
 CC (SBF).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J04761; AAA40519.1; -.
 DR PIR; I54222; I54222.
 DR InterPro; IPR002657; BilAc/Na_sympor.
 DR Pfam; PF01758; SBF; 1.
 KW Transmembrane; Transport; Symport.
 FT NON TER 1 1
 SQ SEQUENCE 182 AA; 19629 MW; 472D732820CDD620 CRC64;

Query Match 9.1%; Score 180.5; DB 1; Length 182;
 Best Local Similarity 30.2%; Pred. No. 4.6e-07;
 Matches 45; Conservative 31; Mismatches 54; Indels 19; Gaps 2;

```

Qy      139 GMMPLCIYLYTWSWSLQQNLTIPTQYQNIQITLVCLTIPVAFGVYVNYRWPQSKIILKIGA 198
          | :||  :|:: | : : | :|  | ||: : ||:| || : : || |::|::
Db      2   GFLPLSSAIYSYLLSIHETLHVPISKILGTLLFIAIPIAAGVVIKSKLPKFSELLQVIK 61

Qy      199 VVGGVLLL-----VVAVAGVVLAKGSWNSDITLLTISFIFPLIGHVTGFLLALEFT 248
          :|||  | : || |  : : | ||:| : | : ||:
Db      62 PFSFILLGGLFLAYHMGVFILVGVRL-----PIVLVGFTVPLVGLLVGYSLAICL 112

Qy      249 HQSWQRCRTISLETGAQNIQMCITMLQLS 277
          : ||:|:| | || : : |||||
Db      113 KLPVAQRRTVSIEVGVQNSLLALAMLQLS 141
  
```

RESULT 10

YCXA_BACSU

```

ID  YCXA_BACSU      STANDARD;      PRT;      409 AA.
AC  Q08791;
DT  01-FEB-1995 (Rel. 31, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Hypothetical protein ycxA (ORF5).
GN  YCXA OR BSU03530.
OS  Bacillus subtilis.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=1423;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168 / JH642;
RX  MEDLINE=93360813; PubMed=8355609;
RA  Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA  Venema G., van Sinderen D.;
RT  "Sequence and analysis of the genetic locus responsible for surfactin
RT  synthesis in Bacillus subtilis.";
RL  Mol. Microbiol. 8:821-831(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE=97124189; PubMed=8969502;
RA  Yamane K., Kumano M., Kurita K.;
RT  "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:
RT  determination of the sequence of a 146 kb segment and identification
RT  of 113 genes.";
RL  Microbiology 142:3047-3056(1996).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE=98044033; PubMed=9384377;
RA  Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA  Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA  Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA  Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA  Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
  
```

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";

RL Nature 390:249-256(1997).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X70356; CAA49820.1; -.

DR EMBL; D50453; BAA08987.1; -.

DR EMBL; Z99105; CAB12147.1; -.

DR PIR; I40489; I40489.

DR SubtiList; BG10172; ycxA.

DR InterPro; IPR007114; MFS.

DR InterPro; IPR005828; Sub_transporter.

DR Pfam; PF00083; sugar_tr; 1.

DR PROSITE; PS50850; MFS; 1.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 9 29 POTENTIAL.

FT TRANSMEM 49 69 POTENTIAL.

FT TRANSMEM 77 97 POTENTIAL.

FT TRANSMEM 100 120 POTENTIAL.

FT TRANSMEM 135 155 POTENTIAL.

FT TRANSMEM 168 188 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 253 273 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.

FT TRANSMEM 309 329 POTENTIAL.

FT TRANSMEM 341 361 POTENTIAL.

FT TRANSMEM 374 394 POTENTIAL.
SO SEQUENCE 409 AA; 44858 MW; 8958A43E87E29DD3 CRC64;

Query Match 6.3%; Score 125; DB 1; Length 409;
Best Local Similarity 20.9%; Pred. No. 0.018;
Matches 78; Conservative 61; Mismatches 122; Indels 112; Gaps 19;

Qy	20	LPVGLVHGNLELVFTVSTVMM--GLLMFSLGCSVE---IRKLWSHIRRPWGIAVGLLC	74
		: : : : : : : : :	
Db	34	LPMADAFHADRSLSVSVSIFMITTGIVQFFVGGFFIDRFSVRKI-----MALGAVC	84
Qy	75	---QFGLMPFTAYLLAISFS'LKPVQAI AVLIMG---CCPGGTISNIFTFWVDGMDLSI	127
		: : : : : :	
Db	85	ISASFLVLPYSPNVHVS-----AIYGVLGIGYSCAVGVTTQYFISCFWFDTHKGLAL	137
Qy	128	SMTTCSTVAALGMMPLCIYL---YTWSWSLQQNLTIPIYQNI GITLVCLTIP-VAFG----	179
		: : : : : : : : : :	
Db	138	AILTNANSAGLVVSPPPITWAAAPYHAGW--QSTYTI----LGIVMAAVLVPLLVFGMKHP	191
Qy	180	-----VYVNYRW-----PKQSKI--ILKIGAVVGGVLLLV-----AV	210
		: : : : :	
Db	192	PHAQAETVKKSYDWRGFWNVMKQSR LIHI LYFGVFTCGFTMGIIDAHLPILKDAHVSHV	251
Qy	211	AGVVLAKGS-----WNSDI-----TLLTISFIFPLIGHVTGFL LALFTHQS--W-	252
		: : : : : : :	
Db	252	NGMMAAFGA FIIIGGLLAGWLSDLLGSR SVMLSILFFIRLLSLICLLIPILGIHHS DLWY	311
Qy	253	-----QRCRTISLETGAQNIQMCITMLQLSFTA EHLVQMLSFPPLAYG	294
		: : : : : :	
Db	312	FGFILLFGLSYTGVIPLTAASISESYQTG-----LIGSLLGINFFI HQVAGALS VYAGGL	366
Qy	295	LFQLIDGFLIVAA	307
		: : :	
Db	367	FEDMTHGYLLIVA	379

RESULT 11

YJIY ECOLI

ID YJIY ECOLI STANDARD; PRT; 721 AA.

AC P39396;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein yjiY.

GN YJIY OR B4354.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC: Enterobacteriaceae; Escherichia.

OX NCBI TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

```
RC STRAIN=K12 / MG1655;
```

RX MEDLINE=95334362; PubMed=7610040;

RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,

RA Blattner F.R.;

RT "Analysis of the Escherichia coli genome VI: DNA sequence of the

RT region from 92.8 through 100 minutes.";

RL Nucleic Acids Res. 23:2105-2119(1995).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -!- SIMILARITY: Belongs to the cstA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U14003; AAA97251.1; -.
 DR EMBL; AE000506; AAC77310.1; -.
 DR PIR; S56580; S56580.
 DR EcoGene; EG12586; yjiY.
 DR InterPro; IPR003706; CstA.
 DR Pfam; PF02554; CstA; 1.
 KW Hypothetical protein; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 169 189 POTENTIAL.
 FT TRANSMEM 197 217 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 263 283 POTENTIAL.
 FT TRANSMEM 288 308 POTENTIAL.
 FT TRANSMEM 332 352 POTENTIAL.
 FT TRANSMEM 380 400 POTENTIAL.
 FT TRANSMEM 469 489 POTENTIAL.
 FT TRANSMEM 491 511 POTENTIAL.
 FT TRANSMEM 537 557 POTENTIAL.
 FT TRANSMEM 574 594 POTENTIAL.
 FT TRANSMEM 602 622 POTENTIAL.
 FT TRANSMEM 674 694 POTENTIAL.
 SQ SEQUENCE 721 AA; 77857 MW; D341DB9C1988C2C8 CRC64;

Query Match 5.9%; Score 117; DB 1; Length 721;
 Best Local Similarity 27.0%; Pred. No. 0.13;
 Matches 55; Conservative 25; Mismatches 66; Indels 58; Gaps 11;

Qy 67 GIAVG--LLCQFGLMPFTAYLLAISFSLKPVQAI AVLI-----MGCCP 107
 | | | | | : | | : | | | | | | | | | | | | | | | | | | |
 Db 107 GPLVGPVLAAQMGYLP GTLWLLAGVVLGAVQDFMVLFISSRRNGASLGEMIKEEMGPVP 166
 Qy 108 GGTISNIFTFWVDGDM DLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI PYQNI GI 167
 | | | : | : | : : : | | | | | | | | | | | | | | | | :
 Db 167 -GTIALFGCFLI---MIIILAVLALIVVKALAES P-----W-----GV 200
 Qy 168 TLVCLTIPVA--FGVYVNYRWP KQSKIILKIGAV-VGGVLLL VVAV--AGVVLAKGSWNS 222
 | | | : | | | : : | | : : | | | | : | | | | | | | | | |
 Db 201 FTV CSTVP IALFMGIYMR FIRPG-----RVGEVSVIGIVLLVASIYFGGVIAHDPYWG P 254
 Qy 223 DITLLTISFIFPLIGHVTGFL LAL 246

Db :| : | |||: |: ||
255 ALTFKDTTITFALIGY--AFVSAL 276

RESULT 12

YCXE_BACME

ID YCXE_BACME STANDARD; PRT; 286 AA.
AC P40419;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 30.5 kDa protein in gdhI 5'region (ORF 2).
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM 1030;
RA Mitamura T., Ehora R.V., Nakai T., Makino Y., Negoro S., Urabe I.,
RA Okada H.;
RT "Structure of isozyme genes of glucose dehydrogenase from Bacillus
RT megaterium IAM1030."
RL J. Ferment. Bioeng. 70:363-369(1990).
CC -!- DEVELOPMENTAL STAGE: Expressed during sporulation.
CC -!- SIMILARITY: TO A SIMILAR ORF IN B.SUBTILIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90043; BAA14098.1; -.
DR PIR; I39851; JS0384.
DR InterPro; IPR004673; RhaT.
DR TIGRFAMS; TIGR00776; RhaT; 1.
KW Hypothetical protein; Sporulation.
SQ SEQUENCE 286 AA; 30490 MW; 95AB89D02511D74D CRC64;

Query Match 5.9%; Score 116; DB 1; Length 286;
Best Local Similarity 21.1%; Pred. No. 0.062;
Matches 64; Conservative 44; Mismatches 91; Indels 104; Gaps 13;

Qy 66 WGIAVGLLCQFGLMPFT-----AYLLAISFSLKPVQAI AVLIMGCCPGGTISNIFTF 117
 || | : | :: | : :| : || | | : : |
Db 13 WGSIVLFNVKLGGGPYSQTLGTTLGALIFSIGIYIFVHPTFTPLIFGV---GVVSGL--F 67

Qy 118 WVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTIPIYQNI GITLVCLTIPVA 177
 | | | : : : ||: : | : || ::
Db 68 WAVGQ---SNQLKSIDLIGVSKTMPI-----STGLQLVSTSL--- 101

Qy 178 FGVYVNYRWPQSKIILKIGAVVGGVLLLVVAVAGVVLA-----KGSWNSDI 224
 ||| | : | :: ||| ||| | : : | : ||| || : : |
Db 102 FGVIVFHEWSTKTSIIL-----GVLALIFIIVGIVLASLQSKKEEKEAEEGKGNFKKGI 154

Qy 225 TLLTISFIFPL-----IGHVTGFLALFTHQSWQRCRTISLETG 263
 :| || : | || || :| || : : : : :
 Db 155 VILLISTVGYLVYVVVARLFNVDGWSALLPQAIGMVIGGVLLTFKHKPFNKYAIRNIIPG 214

Qy 264 AQNIQMCITMLQLSFTAHLVQMLSFP-----LAYGLFQL-----IDGFLIVAAYQTYK 312
 | : | :: :| | : : | | : : | :| :| |
 Db 215 -----LIWAAGNMFLFISQPKVGVATSFSLSQMGIVISTLGGIILGEKKT-K 261

Qy 313 RRL 315
 | : |
 Db 262 RQL 264

RESULT 13

CYB_TOXGO

ID CYB_TOXGO STANDARD; PRT; 368 AA.
 AC O20672; O20928;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB OR CYB.
 OS Toxoplasma gondii.
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
 OC Toxoplasma.
 OX NCBI_TaxID=5811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tournel C., Tomavo S.;
 RT "Cytochrome B of Toxoplasma gondii.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE OF 10-368 FROM N.A.
 RC STRAIN=RH;
 RA McFadden D.C., Boothroyd J.C.;
 RT "Cytochrome B gene from Toxoplasma gondii.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 CC complex (complex III or cytochrome b-c1 complex), which is a
 CC respiratory chain that generates an electrochemical potential
 CC coupled to ATP synthesis (By similarity).
 CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 CC or b566) is high-potential and absorbs at about 566 (By
 CC similarity).
 CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 CC cytochrome c1 and the Rieske protein (By similarity).
 CC -!- SIMILARITY: Belongs to the cytochrome b family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF015627; AAB82741.1; -.
 DR EMBL; AF023246; AAC34138.1; -.
 DR InterPro; IPR005798; Cytb_b6_C.
 DR InterPro; IPR005797; Cytb_b6_N.
 DR Pfam; PF00032; cytochrome_b_C; 1.
 DR Pfam; PF00033; cytochrome_b_N; 1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KW Heme.
 FT METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 178 178 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 192 192 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 368 AA; 41594 MW; CC7C6BD3784287CA CRC64;

Query Match 5.7%; Score 112.5; DB 1; Length 368;
 Best Local Similarity 24.6%; Pred. No. 0.15;
 Matches 57; Conservative 37; Mismatches 101; Indels 37; Gaps 8;

Qy 84 YLLAISFSLKPVQAIIVLIM---GCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALG 139
 :|:|:| | : : | : | : : : | : | | |
 Db 34 FLVAMTFVLQIITGITLAFRYTSEASCAFASVQHLVREVAAGWEFRMLHATTASFVFLCI 93
 Qy 140 MMPLCIYLYTWSWSLQQNLTIPTYPQNIIGITLVCLTIPVAFGVYVNYRWPQSKIIKIGAV 199
 :: : || ||:| || : : |:| ||| || || | : | ||
 Db 94 LIHMTRGLYNWSYSY---LTTAWMS-GLVLYLLTIATAFLGYV-LPWGQMS----FWGAT 144
 Qy 200 VGGVLLLVAAGVVLAKGSWNSDITL---LTISFIFPLIGHVTGFLALFTHQSWQRCR 256
 | || : | | : ||:| : || | || : | : | :
 Db 145 VITNLLSPIPYLVPWLLGGYYVSDVTLKRFFVLHFIPLFIGCIIIVLHIFYLHLN----- 199
 Qy 257 TISLETGAQNIQMCITMLQLSF-----TAEHLVQMLSFPPLAYGLFQL 298
 | : | | | : : | : : : : | : | : |
 Db 200 -----GSSNPAGIDTALKVAFYPHMLMTDAKCLSYLIGLIFLQAAFLMEL 245

RESULT 14

Y944_SYNY3

ID Y944_SYNY3 STANDARD; PRT; 383 AA.

AC P74311;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein slr0944.

GN SLR0944.

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,


```

RA      Tabata S.;
RT      "Sequence analysis of the genome of the unicellular cyanobacterium
RT      Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT      entire genome and assignment of potential protein-coding regions.";
RL      DNA Res. 3:109-136(1996).
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC      -!- SIMILARITY: BELONGS TO THE ACR3 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D90914; BAA18405.1; -.
DR      PIR; S76146; S76146.
DR      InterPro; IPR004706; Acr3.
DR      InterPro; IPR002657; BilAc/Na_symport.
DR      Pfam; PF01758; SBF; 1.
DR      TIGRFAMS; TIGR00832; acr3; 1.
KW      Hypothetical protein; Transmembrane; Complete proteome.
FT      TRANSMEM      25      45      POTENTIAL.
FT      TRANSMEM      53      73      POTENTIAL.
FT      TRANSMEM     103     123      POTENTIAL.
FT      TRANSMEM     139     159      POTENTIAL.
FT      TRANSMEM     166     186      POTENTIAL.
FT      TRANSMEM     200     220      POTENTIAL.
FT      TRANSMEM     238     258      POTENTIAL.
FT      TRANSMEM     272     292      POTENTIAL.
FT      TRANSMEM     309     329      POTENTIAL.
FT      TRANSMEM     332     352      POTENTIAL.
SQ      SEQUENCE      383 AA;  42402 MW;  3D8C4CF8EA2FF08B CRC64;

```

Db 295 LKLKL-----SYEDAAPAALIGASNHFEVAIATAVMLF 327

RESULT 15

NU5M_ANOQU

ID NU5M_ANOQU STANDARD; PRT; 576 AA.
AC P33510;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5.
OS Anopheles quadrimaculatus (Mosquito).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=7166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Orlando;
RX MEDLINE=92190510; PubMed=2134168;
RA Cockburn A.F., Mitchell S.E., Seawright J.A.;
RT "Cloning of the mitochondrial genome of Anopheles quadrimaculatus.";
RL Arch. Insect Biochem. Physiol. 14:31-36(1990).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04272; AAA93547.1; -.
DR InterPro; IPR003916; NADHub_oxred5.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR001516; Oxidored_q1_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF00662; oxidored_q1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 576 AA; 65913 MW; A82B45D67F430F42 CRC64;

Query Match 5.4%; Score 107.5; DB 1; Length 576;
Best Local Similarity 21.6%; Pred. No. 0.56;
Matches 80; Conservative 52; Mismatches 105; Indels 133; Gaps 21;

Qy 8 SSACPANSSEELPVGLEVHGN-----LELVFTVVST-----VMMGLLMFSLG 50
|| || : || || : | : | : | : || || |
Db 200 SSWLPA-AMAAPTIVSALVHSSTLVTAGLYLLIRENILLTDWWMGQFMLLISGLTMFMAG 258

Qy 51 CSV----EIRKLWSHIRRPWGIAVGLLCQFGLM----PFTAYLLAISFSLKPVQAI AVLI 102
::|: ||: | | || | || | | : |
Db 259 LGANFEFDLKKI-----IALSTLSQLGLMMSILSMGFYKLAFFHLLTHALFKALLF 309

Qy 103 MGCCPGGTISNIFT---FWVDGDMDLISMT-TCSTVAALGM--MPLCIYLYTWSWSLQQ 156

	:	: : : : : :	:	:
Db	310	M--CAGSIIHNMKNSQDIRMMGSLSMSPLTCSCFNVANLALCGMPFLAGFYSKDLILEM	367	
Qy	157	NLTIPYQNI-----GITLVCLTIPVAFGVYVNYRWPKQSKII-----	193	
		::: :	: : : :	::
Db	368	-VLSYVNVFSFFLFFFSTGLT-VCYSFRL---VYYSMTGDFNSSVLHPLNDSGWTMLFS	422	
Qy	194	---LKIGAVVGGVLLLVAAGVVLAKGSW-----NSDITLLTISFIFPLIGHV	239	
		: :		: : : : : :
Db	423	IFFLMIMAVIGGSML-----SWLMFLNPSMICLPFDLKMLTL-FVC-ILGGL	467	
Qy	240	TGFLALFTHQSWQRCRTISLETGAQNIQMCITMLQLSF-----TAEHLVQM	286	
		: :	: :	: :
Db	468	IGYLLS-----NVSLEFTNKALYFYNFYFAGSMWFMPVVSTIGV	507	
Qy	287	LSFPLAYGLF	296	
		::: :		
Db	508	INYPLKLGLY	517	

Search completed: March 23, 2004, 14:36:18
Job time : 19 secs